

© 2017 Katrina S. Fenlon

THEMATIC RESEARCH COLLECTIONS: LIBRARIES AND THE EVOLUTION OF
ALTERNATIVE SCHOLARLY PUBLISHING IN THE HUMANITIES

BY

KATRINA S. FENLON

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Library and Information Science
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2017

Urbana, Illinois

Doctoral committee:

Professor Carole Palmer, University of Washington, Chair and Director of Research
Senior Lecturer Maria Bonn
Professor Julia Flanders, Northeastern University
Professor Allen Renear

ABSTRACT

Scholarship across disciplines is changing in the face of digital methodologies, novel forms of evidence, and new communication technologies. In the humanities, scholars are confronting and often pioneering innovative modes of viewing, reading, interacting with, collecting, interpreting, contextualizing, and sharing their sources and derived evidence. From research blogs to multimedia products to large-scale digital corpora, new forms of scholarly production challenge conventions of publishing and scholarly evaluation and the long-term maintenance of scholarship in libraries. The omission of digital scholarship from systems of scholarly communication – including peer review, discovery, organization, and preservation – poses a potential detriment to the evolution of humanities scholarship and the completeness of the scholarly record.

One emergent genre of digital production in the humanities is the *thematic research collection* (Palmer, 2004): a collection of primary sources created by scholarly effort to support research on a theme. Thematic research collections constitute a diverse genre with a range of functions beyond supporting research: collections serve as hubs for experimentation, collaboration, and communication; facilitate the reuse of humanities data; generate new lines of inquiry and original evidence; and engage broad audiences. Yet, despite their significant and distinctive contributions to scholarship, thematic research collections have struggled to gain integration into systems of evaluation and post-publication management in libraries, in part because we do not know enough about them.

This study investigates the defining features of thematic research collections and considers the challenges for libraries in supporting this genre. Through a typological analysis of a large sample of collections in tandem with a qualitative content analysis of representative collections, this study identifies different types of thematic research collections, which make different kinds of contributions to scholarship. Through interviews with practitioners in digital humanities centers and libraries, this study illuminates challenges to the sustainability and preservation of thematic research collections, and potential strategies for ensuring their long-lived contributions to scholarship. This study lays a foundation for understanding collections as a significant, dynamic, vibrant exemplar of how digital scholarship continues to evolve, with implications for library practice and the evolution of research and communication across disciplines.

ACKNOWLEDGEMENTS

I am profoundly indebted to my advisor and director of research, Carole Palmer, for years of guidance, instruction, editing, and mentorship. I will always strive to reach the level of scholarship that Dr. Palmer has modeled for me.

I extend my gratitude to my esteemed committee. I cannot imagine a more brilliant, thoughtful, and generous council.

I am grateful for constant aid – material and otherwise – from the School of Information Sciences, its vigilant staff, and its seemingly tireless administration. This dissertation work was partially supported by the Josie B. Houchens Fellowship.

My work has been inspired and improved by guidance from and collaborations with many co-adventurers in research over the years, including Jacob Jett, Megan Senseney, Dr. Karen Wickett, Timothy Cole, Dr. Stephen Downie, and many others.

I doubt I would have finished this project without having discovered a second home/third space just when I needed it – Seven Saints, run by my friend, the peerless Anne Clark. I am inexpressibly glad for the community I found there, and for so many surprising opportunities for joy and growth.

I am grateful to Dr. John Jones, for helping me see the beauty in the process.

Finally, I am especially thankful to my family and dear friends for the generous gifts they have given me in pursuit of this goal: the repeated assurances of loving support, possibilities for free and unburdened time, rambling phone calls, offers to read my delirious drafts, nights of quiet solidarity, nights of raucous solidarity, and lots of patient help finding my keys. I am forever grateful to my brilliant dissesister, Andrea Thomer, who has proven a consummate ally on this journey. I extend my deepest thanks of all to my perspicacious and lionhearted mother, Evelyn Fenlon; my sister, Alison Fenlon, with her iridescent mind and capacious soul; and my dearest Noah Dibert, who is my daily angel of compassion, aspiration, and true grit.

TABLE OF CONTENTS

CHAPTER 1: INTRODUCTION	1
1.1. PROBLEM SPACE	1
1.2. THEMATIC RESEARCH COLLECTIONS	3
1.3. RESEARCH QUESTIONS AND SUMMARY OF APPROACHES	6
1.4. CONTRIBUTIONS	7
CHAPTER 2: LITERATURE REVIEW	8
2.1. SCHOLARLY COMMUNICATION: A SHIFTING LANDSCAPE	8
2.2. COLLECTIONS GENERALLY	19
2.3. THEMATIC RESEARCH COLLECTIONS	23
CHAPTER 3: METHODS.....	33
3.1. OVERVIEW OF APPROACHES	33
3.2. PROVISIONAL TYPOLOGY OF COLLECTIONS.	35
3.3. QUALITATIVE CONTENT ANALYSIS	40
3.4. INTERVIEWS.....	47
CHAPTER 4: COLLECTION PURPOSES	52
4.1. INTRODUCTION	52
4.2. FOUNDATIONAL PURPOSES	54
4.3. GENERATIVITY.....	59
4.4. AUDIENCES	65
CHAPTER 5: KINDS OF COLLECTIONS.....	76
5.1. INTRODUCTION	76
5.2. PROVISIONAL TYPOLOGY	77
5.3. ENRICHING TYPOLOGY WITH CONTENT AND CONTEXT	81
5.4. COMPLETENESS.....	86
5.5. PROPOSED KINDS OF COLLECTIONS	90
CHAPTER 6: SUSTAINABILITY AND PRESERVATION	107
6.1. CHALLENGES	108
6.2. STRATEGIES	114
6.3. ROLES.....	120
6.4. EXTENDING CURRENT FRAMEWORKS	124
CHAPTER 7: COLLECTIONS AS PLATFORMS	132
7.1. CHALLENGES AND OPPORTUNITIES FOR LIBRARIES	132
7.2. DEFINING FEATURES OF COLLECTIONS	139
7.3. CONCLUSIONS AND FUTURE WORK.....	147

REFERENCES.....	149
APPENDIX A: TYPOLOGY OF THEMATIC RESEARCH COLLECTIONS	171
APPENDIX B: CONTENT ANALYSIS PROTOCOL	177
APPENDIX C: INTERVIEW PROTOCOL.....	191

CHAPTER 1: INTRODUCTION

1.1. PROBLEM SPACE

Changes in scholarly communication and publishing over the past couple decades have yielded new kinds of research products in the humanities,¹ ranging from blogs, to multimedia resources that function as hubs for discourse communities, to digital scholarly editions, to massive textual corpora. Beyond changing economic and technical models for digitally publishing genres of research that are familiar from our printed history (such as digital scholarly editions, digital monographs, or electronic journals), evolutions in digital scholarship have produced less familiar varieties of publication, born from technologically enabled changes in how humanities research is conducted, in the nature of historical and literary evidence, and in what scholars are able and want to share.

One genre of digital production in the humanities is the *thematic research collection* (Palmer, 2004; Unsworth, 2000), which has been defined as a collection, created by scholarly work, which presents primary source evidence and related materials in order to support research on a theme (Palmer, 2004). For more than a decade, the thematic research collection has been acknowledged as a genre of scholarly production (see for example Unsworth, 2000; Brockman et al., 2001; Alonso et al., 2003; Palmer, 2004; Schreibman et al., 2008; Ciula & Lopez, 2009; Price, 2009; Flanders, 2014; Thomas, 2015).

Alternative scholarly products, including the thematic research collection, stand largely outside of established systems of publication and library collection. Certain points on the cycle of scholarly communication (see Figure 1) raise barriers to the immediate discoverability and long-term usefulness of alternative scholarly products. Scholars have struggled to find venues for their review. Dissemination is often just putting a resource on the Web, without the scaffolding of library or publisher support. Provisions may not be made for centralized discovery and access, or long-term access to these resources, as they are not normally treated as part of a research library collection (and are rarely indexed elsewhere). It is the argument of this project that the omission

¹ These changes are not limited to the humanities. In the sciences, emergent kinds of shared products include openly accessible data sets, intermittent and informal publication of research results, open peer review for intermittent findings, publication of software tools, etc. In the humanities, we see roughly similar things: humanities data sets, intermittent publication or informal sharing of research threads through a wide array of venues, publication of software tools, digital scholarly editions that are both documentary and critical (Flanders, 2014), annotations, etc.

of innovative digital products from systems of publication and library collection poses a potential detriment to the evolution of humanities scholarship and the completeness of the scholarly record.



Figure 1. "Publication cycle"²

For these reasons, our existing understanding of thematic research collections, among other alternative forms, is inadequate to leverage their full value. They are recognized by the research community as scholarly products, but without common systems for publication and evaluation, scholars struggle to obtain reliable support to publish and get credit for diverse contributions, hobbling the kind of research production scholars want to do. Without common systems for the integration of new scholarly products into our libraries, we compromise the immediate discoverability and accessibility of these new products, and the completeness of the scholarly record over time. Library systems for the description, discovery, and maintenance of scholarship have fallen behind evolutions of digital scholarship in the humanities. Our lack of knowledge about

² From <<https://library.uwinnipeg.ca/scholarly-communication/index.html>>. Consider also, from the LIS perspective, what Tennis (2011) describes as a “five-stage cycle”: creation, publication, organization, access, and preservation, which cycle “constitutes the core concern for much of library and information science.” In that cycle, alternative scholarly products suffer most in the stages of organization and preservation.

the nature and roles of alternative publications in scholarly communication, along with services for their publication and ongoing discovery, access, and use, together pose a significant potential impediment to their ongoing usefulness as scholarly products. Thomas (2016) has called on humanities scholars to examine, discuss, and clarify new genres, including thematic research collections, so that we may understand how to characterize and evaluate their contributions. A better understanding will also help us improve services and support to authors and users of new scholarly products.

This project aims to deepen our understanding of the genre of thematic research collections, including their defining features, their commonalities and how they are different, both from one another and from other kinds of collections. Through a provisional typology of a broad base of thematic research collections, augmented with a close content analysis of exemplars of “types,” this project investigates the nature and roles of thematic research collections in scholarly communication. The project follows on that empirical study of thematic collections with a study of their library contexts. A set of interviews with professionals will investigate current practice around thematic research collections in digital humanities centers and libraries, particularly considering the challenges and opportunities that confront the long-term service to and support of this genre.

1.2. THEMATIC RESEARCH COLLECTIONS

For more than a decade, the (digital) thematic research collection has been acknowledged as a genre of scholarly production in the humanities (Unsworth, 2000; Brockman et al., 2001; Alonso et al., 2003; Palmer, 2004; Schreibman et al., 2008; Ciula & Lopez, 2009; Price, 2009; Flanders, 2014; Meiman, 2015; Thomas, 2016). A thematic research collection is a digital collection, created by scholarly work, which aggregates primary source evidence and related materials, in order to support research on a theme (Palmer, 2004). This kind of collection occupies a liminal space, functioning both as a platform for research that is leveled upon primary sources, and simultaneously as a “presentation” or publication of scholarly work.

Palmer (2004) elaborates on Unsworth’s original list of characteristics of thematic collections (Unsworth, 2000), as depicted in Figure 2.

Content	Function
Basic elements	
* Digital	Research support
* Thematic	
Variable characteristics	
* Coherent	Scholarly contribution
* Heterogeneous	Contextual mass
* Structured	Interdisciplinary platform
* Open-ended	Activity support

Figure 2. Palmer's (2004) Features of thematic research collections

Canonical exemplars of thematic research collections include the William Blake Archive,³ the Dickinson Electronic Archives,⁴ the Walt Whitman Archive,⁵ and Valley of the Shadow.⁶ Those who coined and first characterized the term “thematic research collection” in the early 2000s did so in recognition of these exemplars, which continue in active development and use by humanities scholars.

Thematic research collections are heterogeneous, both internally and among themselves:

- Within: thematic research collections are internally heterogeneous, perhaps more so than other genres of scholarly product. They are often multimedia endeavors. They “collocate” sources in ways that are common on the web but uncommon in traditional approaches to either collection or research publication (for example, using a mix of hyperlinking and embedding). They juxtapose and even blend varieties of primary and secondary sources, layering and linking evidence and interpretation. As Palmer (2004) notes:

The capabilities of networked, digital technology make it possible to bring together extensive corpuses of primary materials and to combine those with any number of related works. Thus the content is *heterogeneous* in the mix of primary, secondary, and tertiary materials provided, which might include manuscripts, letters, critical essays, reviews, biographies, bibliographies, etc., but the materials also tend to be multimedia.

- Between: thematic research collections are different one from another. We have suggested their great range in purpose, form, and function. The first phase of this study analyzes this range in more detail.

³ <<http://www.blakearchive.org/blake/>>

⁴ <<http://www.emilydickinson.org/>>

⁵ <<http://whitmanarchive.org/>>.

⁶ <<http://valley.lib.virginia.edu/>>

By way of illustration of their heterogeneity, consider briefly a juxtaposition of two collections. Many thematic research collections take the form of a digital archive. The Rossetti archive, for example, “facilitates the scholarly study of Dante Gabriel Rossetti,” 19th-century painter, designer, writer, and translator.⁷ Much like an archive, it “provides access to all of [Rossetti’s] pictorial and textual works and to a large corpus of contextual materials,” including “high-quality digital images of every surviving documentary state” of the works.” The works are encoded and fully searchable, and primary sources are “transacted with substantial body of editorial commentary, notes, and glosses” – but primary and secondary are clearly distinguished.⁸

The Rossetti Archive is a traditional thematic research collection, often cited in the literature on the genre and in (digital) humanities literature generally. But such labels as “traditional” and “archive” belie the fundamental similarity between this project and other more experimental projects that meet our definition. Consider “O Say Can You See: Early Washington, D.C., Law, & Family.”⁹ This resource is a “deep relationship mapping,” or network, of people in early Washington, D.C. This network is derived from a collection of case files and kinship and family records, which the site also makes readily available and searchable. At heart, the collection is a conventional collection of primary sources, but it functions like a layer on top of one. It is born of analytic work, and offers novel productions (both the collection of data itself and the network or mapping derived from that data). It is the product of research (and therefore indisputably a scholarly production), but with the intent to facilitate research in the way that a simpler collection of primary sources does.

Some thematic research collections cater more to research, others more to pedagogy, and many to both. While some make little or no explicit argumentation, others are more discursive: alongside the primary sources they offer coherent arguments, or narrative interpretation. For some conventional search and browse constitute the primary mode of interaction; for others this mode is secondary or (maybe) nonexistent. Many cases must be considered “edge cases,” which conform only questionably to our existing definition of thematic research collections. These edge cases may nonetheless prove to be central to this study, as they shed light on the diversity of the genre, and expand our conception of what falls in it.

⁷ “The complete writings and pictures of Dante Gabriel Rossetti: A hypermedia archive”

<<http://www.rossettiarchive.org/>>

⁸ <<http://www.rossettiarchive.org/>>

⁹ <<http://earlywashingtondc.org/>>

Thematic research collections pose a ripe subject for study within the problem space articulated above, as one example of a new, vibrant, diverse digital genre that is underserved by existing publication and collection systems.

1.3. RESEARCH QUESTIONS AND SUMMARY OF APPROACHES

This project addresses the following research questions.

- (R1) What are the defining features of thematic research collections as a scholarly genre?
 - What features are common to thematic research collections?
 - What features distinguish thematic research collections from other kinds of collections?
 - What kinds of thematic research collections are there, and how are they distinguished from one another?
- (R2) What are the challenges, for libraries and related scholarly-publishing entities, in supporting thematic research collections as a scholarly genre?
 - How do library publishing programs and related scholarly-publishing entities support the creation and publication of thematic research collections, and what problems exist in meeting the needs of collection creators?
 - How do libraries collect, represent, describe, preserve, and otherwise treat thematic research collections after publication, and what problems exist in meeting user needs?

A provisional typology of a broad base of thematic research collections, augmented with a close content analysis of exemplars of “types” is used investigate (R1). The typological work aims to evoke the full range of the genre, as it is defined, along with a set of potential defining features. It provides a foundation for a content analysis of exemplary collections, selected to represent diverse types within the genre, which explores the commonalities and differences in more detail and identifies some defining characteristics of thematic research collections. Interviews with representatives of digital humanities centers and libraries were conducted to identify current practice around thematic research collections and reveal challenges to and potential strategies for their integration into library systems of collection, discovery, access, and ongoing maintenance. This phase of the project addresses (R2).

1.4. CONTRIBUTIONS

This project aims to contribute to our understanding of how libraries may continue to cultivate and curate new forms of digital humanities scholarship. This study affords a set of defining characteristics of thematic research collections, an investigation of how those characteristics are manifested by collection design to support different kinds of contributions to scholarship, and substantive leads on the challenges confronting the sustainability of the genre. The results lay a foundation for further research into how libraries can more systematically integrate these resources into existing collection/access infrastructures, to ensure their ongoing discoverability and usefulness. The shapes that thematic research collections take in order to serve their multifaceted purposes and diverse audiences offer new directions for leveraging humanities evidence scattered across the Web. This study hopes to lay a foundation for understanding collections as an especially interesting exemplar of how digital scholarship continues to evolve, with implications for the evolution of library practice.

Audiences likely to have some interest or stake in the outcomes of this work include the growing library publishing community; the digital humanities community; practitioners of collection development and description, especially those with an interest in standards-development; researchers interested in scholarly communication generally, including from an information-behavior or use-and-users perspective; and humanities scholars engaged with digital collections of primary sources, either as builders or users.

CHAPTER 2: LITERATURE REVIEW

This section reviews the literature on several facets relevant to this research. I begin by contextualizing thematic research collections within the shifting landscape of scholarly communication, which has disrupted traditional institutional roles in publication, systems of evaluation, and posed new challenges to the processes entailed in the activity of library collection. I follow that contextual exploration with an examination of collections in a more general sense, because our understanding of thematic research collections in library and information science has been contextualized by conceptual accounts of other kinds of collections, and studies of the roles of collection in humanities scholarly practice. Finally, this review explores existing characterizations of thematic research collections, their nature, and their evolving place in humanities discourse.

Pieces of this literature review have appeared in other published and unpublished works of mine: parts of section 2.1 draw on a literature review I wrote as an appendix to a proposal to the Andrew W. Mellon Foundation, “Publishing Without Walls: Understanding the Needs of Scholars in a Contemporary Publishing Environment” (2015). Section 2.2 draws some from my written field exam in information organization and access (2013). Elements of sections 2.2 and 2.3 appear in Fenlon et al. (2014), which describes a study of humanities scholars’ creation and use of collections, conducted under the province of the HathiTrust Research Center’s “Workset Creation for Scholarly Analysis: Prototyping Project” (2014).

2.1. SCHOLARLY COMMUNICATION: A SHIFTING LANDSCAPE

Throughout the research universe, scholars are leaning towards alternative publications and new modes of dissemination (Palmer, 2005; Thomas, 2015). This trend has complex origins, and while it is not within scope of this literature review to comprehend the full history, we can assert that the trend toward new modes of publication has been enabled by the advent of digital technologies, and has been related to what is widely perceived to be a long-term crisis¹⁰ in scholarly communication (Alonso et al., 2003; Unsworth, 2003; McCormick et al., 2015; and many others). Brown et al. (2007) anticipated a transformation in scholarly publishing that would entail

¹⁰ A “crisis” is widely but not universally perceived. Harley et al. (2010a) show that humanities scholars in certain fields do not perceive a crisis at all.

the creation of fundamentally new publishing formats, along with a sustainable marketplace of highly diversified distribution channels attuned to different kinds of content and audiences.

How are patterns of scholarly communication in the humanities changing in light of the challenges and opportunities of digital publishing? Weller (2011) notes that the “mechanisms through which scholars publish and communicate their findings and learn about the work of others are undergoing radical change.” Yet Brown et al. (2007) observe that scholarly publishing has lagged behind changes in the information consumption patterns of scholars, resulting in an explosion of grey literature and blurring the line between formal and information publication. For example, there is growing evidence of the use of social media and Web 2.0 tools for scholarly communication at all phases of research, including “personal knowledge publishing” (Aimeur et al., 2005). Procter et al. (2010) argue that Web 2.0 provides a technical platform for the “re-evolution” of scientific and scholarly communication. However, assertions of the value of social media for scholarly communication continue to be predominantly speculative rather than empirical (Acord & Harley, 2013).

Despite technologically enabled changes in scholarly communication, there remains a gap between new forms of communication and publication and actual scholarly practice (Cohen, 2013; Harley 2013), attributable to many causes. Humanists remain devoted to long-established channels of dissemination while increasingly employing and even producing new tools, technologies, and diverse information resources in their work (Bulger et al., 2011). Harley (2010b) finds that while scholars have embraced digital primary sources in their research processes, they remain conservative in their own digital dissemination behavior. Nonetheless, experiments in new genres are “taking place within the context of relatively conservative value and reward systems.” Bulger et al. (2011) identify as barriers to new modes of communication, “lack of awareness and institutional training and support, but also lack of standardization and inconsistencies in quality and functionality across different resources.”

Harley et al. (2010a) determined that scholars have a variety of competing criteria for choosing modes of publication. Time to publication is not a primary concern for humanities scholars (as it is for those in the sciences). Scholars compromise between targeting niche audiences and pursuing publications that may have a more general audience. Scholars want to be able to link their publications to primary sources or include media in new ways. They also express an interest in new publishing models for shorter monographs in the humanities. Scholars perceive limitations

in the current publishing system and express growing interest in “the potential of electronic publication to extend the usefulness and depth of final publications,” which may include embedded media. However, none of the scholars interviewed by Harley et al. could identify easy-to-use tools for publishing multimedia monographs. Lack of tools, and lack of institutional support and expertise, pose major barriers to scholars’ engagement with new modes of publication.

Yet interest thrives in alternative forms of publication and scholarly communication in the humanities. Harley et al. (2010a) conducted an extensive set of interviews with scholars across research institutions in seven fields (archeology, astrophysics, biology, economics, history, music, and political science) to analyze how faculty, as a primary stakeholder, value traditional and emerging forms of scholarly communication. They identify a number of faculty values, behaviors, and requirements that bear on new digital publishing initiatives: As others have noted, the conventions of systems of scholarly evaluation remain a primary obstacle to the growth of digital, experimental, and open-access publishing. Nonetheless, humanities departments are increasingly implementing changes to tenure and promotion criteria to embrace new, multimedia, dynamic forms of publication as analogous to print monographs. See, for example, the University of Florida’s guide to “DH in Tenure and Promotion (Digital Humanities Working Group, University of Florida), Modern Language Association (2012), and AHA Ad Hoc Committee on the Evaluation of Digital Scholarship by Historians (2015).

Accounts of the specific features and characteristics of emergent genres of scholarly publication are largely speculative (with some exceptions; for example, Unsworth, 2000, and Palmer, 2004, note the emergence of the thematic research collection as a genre of scholarly product, and Jewell, 2009, describes the digital scholarly edition). Brown and Simpson (2014) offer a vision of new modes of text, both primary and secondary sources (such as monographs) in the humanities. They assert that texts that humanities scholars produce, publish, and use are increasingly “dynamic, increasingly collaborative, granulated and distributed, and interdependent with other text or data.” Similarly, Ciula and Lopez (2009) assert that humanities scholars want to publish in more creative modes, so that the presentation of their texts reflects their methods of interpretation more fully than in a traditional print monograph. They increasingly want to incorporate primary sources into their publications in dynamic ways, so that texts serve as “connective structures” between resources. Indeed, Weller (2011) asserts that a demand for innovative modes of publication is a primary benefit of open access publishing. Weller notes that

most of the advantages of open access publishing, from the scholarly perspective (such as skirting time lag to publication, or that evaluative metrics are less biased to accessibility over quality), could theoretically be accommodated by the current system of scholarly publishing. However, Weller notes that people are seeking alternative methods of communication, publishing, and debate in new media, and this transcends existing genres and systems of publication.

2.1.1. Shifting institutional roles in scholarly publishing

This section reviews shifting institutional roles in scholarly publishing, prioritizing those that contextualize the development and maintenance of thematic research collections. In particular, this section considers the growth of library publishing, especially for alternative scholarly products, where library publishing is defined broadly as “the set of activities led by college and university libraries to support the creation, dissemination, and curation of scholarly works” (Watkinson et al., 2012).

Libraries increasingly provide publishing services to authors seeking to produce alternative kinds of publication, such as thematic research collections. Thematic research collections are also spawned by collaborations between authors and digital humanities centers, or digital scholarship services. These units or institutions are often related to the library – whether organizationally subsumed, affiliated, or physically co-located (Hahn, 2008; Tracy, 2016; Vandegrift, 2012; Vandegrift and Varner, 2013). Publishing services within libraries are sometimes integrated with other library (or library-based) services and initiatives, including digitization, digital humanities, digital repositories, digital preservation, etc. (Hahn, 2008). Rarely, but increasingly, university presses have begun to pursue alternative genres of publication, sometimes in collaboration with the library.¹¹ As Sundaram (2016) notes in a description of Stanford University Press’ new (Mellon-funded) foray into digital publishing, that program will leave the aspect of development assistance for digital publications to the library:

Not only do the impressive efforts underway at academic libraries around the country show that other players on the academic field are already there to assist authors, we also firmly believe that the process of building digital projects is inherent to the author’s creative process, it is part of the ‘writing’ of digital

¹¹ It is for these reasons that library publishing programs and digital humanities centers are a selected for study in the last phase of this project.

communication, and we as academic publishers should not create but rather edit, produce, and market it. –Sundaram, 2016

In light of digital information technologies and digital publishing, university libraries are expanding their missions to encompass digital publishing, and exploring and supporting new models of scholarly communication. The *Library Publishing Directory 2015* of the Library Publishing Coalition lists 124 library publishers (Lippincott, 2015), up from 35 in 2008 (Hahn et al., 2009, as noted in Bonn and Furlough, eds., 2015).¹² Courant and Jones (2015) argue that libraries “are natural and efficient loci for scholarly publication.” Lefevre and Huwe (2013) even assert that the act of digital publishing is a new core competency for the library profession. Five years ago, Adema and Schmidt (2010) reviewed library involvement in scholarly publishing and found libraries engaged in a limited number of publishing roles, ranging from creating institutional repositories and supporting new scholarly communication activities to publishing digital, open access journals and books (predominantly in STEM fields and predominantly journals, at that time), and funding author fees where they are required in certain open access publications.

The phenomenon of library publishing has grown significantly in the intervening years. Hahn (2008) finds that the development of publishing services in the library are “being driven by campus demand...Scholars and researchers are taking their unmet needs to the library.” Bonn and Furlough (Eds., 2015) offer a collection of studies highlighting the diversity of existing library publishing programs and services. They see a distinctive niche for libraries in the scholarly publishing ecosystem, in “keep[ing] alive the specialized but commercially unviable works that publishers have increasingly let slip from their lists. Ideally, they can also bring to life new subjects and new formats, including formats of varying length and composition, that have been shunned by traditional publishers.” Scanning the landscape of library publishing, they offer a sort of taxonomy of library publishing (or publishing-related) activities, which range broadly in scope, and which include:

- Digitization of library holdings (often coupled with print-on-demand)
- Original publishing, sometimes through fully fledged imprints: they note that this activity is sometimes organized such that the library absorbs the university press as a unit –

¹² The Mellon Foundation’s program for capacity-building projects for university presses recognizes shifts in the roles and responsibilities for campus publishing by requiring presses to collaborate with other units on campus (Straumsheim, 2015). See <http://www.aaupnet.org/aaup-members/news-from-the-membership/collaborative-publishing-initiatives>

“library-press integration” – with varying levels of dependence, involvement, or control accorded to each institution.

- Forging partnerships with external or internal entities, such as scholarly societies or specific campus departments, to publish specific works.
- Publication of (or curation, management, and provision of access to) humanities and social science data
- Provision of publishing support services: from hosting and distribution (e.g., through institutional repositories), to education and consultation on where and how to publish, and on publishing agreements

The volume also covers libraries’ involvement in educational publishing (e.g., of open-access textbooks), and faculty and student self-publishing. On the subject of monograph publication, Courant and Jones (2015), in Bonn and Furlough (Eds., 2015), consider the economic viability of library publication of open-access monographs and find that the “cost of producing a well-reviewed and lightly edited scholarly monograph to be distributed digitally through libraries” is unlikely to be prohibitive. In the same volume, McCormick (2015) considers creative approaches to publishing in libraries, particularly in employing open-source tools to publish the multimedia output of new scholarly methods.

In light of this trend toward library-based publishing, the boundaries between the activities of university libraries and university presses have become less distinct. Opportunities for partnership are ripe, and offer the academy the potential for increased control over intellectual products (Crow, 2009). In their final report on a survey of the state of library publishing as a whole, Mullins et al. (2012) advocate for further development and professionalization of library publishing roles. They offer a series of recommendations for libraries, which includes leveraging partnerships with university presses to expand from simply hosting digital content to providing more holistic services.

Witnessing a recent increase in experimental, collaborative efforts to enable and explore open access book publishing in the humanities and social sciences, Adema and Schmidt (2010) assert that library/press collaborations in open access monograph publishing offer a solution to the scholarly publication crisis described above. They review several cases that exemplify how collaborations exploit the core competencies of both institutions:

- The University of California Publishing Services provides hybrid publication services for monographic publishing and marketing. In this case, the libraries are responsible for open-access digital publishing, peer review, and management tools, while the press handles sales, distribution, marketing, and print-on-demand. A third partner, Campus Publishing, takes care of content selection, peer review, editing, design, and composition.
- The Newfound Press is a digital imprint from within the University of Tennessee libraries, which publishes peer-reviewed, open-access books and journals. While independent of the university press, it does offer options for print-on-demand through the press.
- Through the Scholarly Publication Office at the University of Michigan, the University of Michigan Library partners with the University of Michigan Press and with the Open Humanities press to publish open monograph series in an academic-led endeavor to experiment with library publishing. The library also launched MPublishing, which unit incorporates all academic publishing activities of the library and expands services from the humanities and social sciences into other areas, such as the biomedical and medical disciplines (Adema & Schmidt, 2010).

A SPARC guide to critical issues for campus publishing partnerships (Crow, 2009) asserts that a transition to long-term, programmatic collaboration between university libraries and presses will require a high degree of interdependence between the institutions. Specifically, they will need to establish administrative and funding structures that integrate without disrupting the disparate competencies of those two institutions, and identify objectives and services according to current and anticipated requirements of faculty and researchers. The guide offers an overview of existing collaborations. Two thirds of collaborations involve just the university library and press; the remaining third include other partners, including computing centers, departments, and societies.

From the contrasting perspective of university presses, Withey et al. (2011) explore sustainable business models for university presses and reiterate the evident potential for beneficial collaboration with libraries, noting that many university presses have partnered with libraries to host open-access digital books. They assert that “[p]artnerships with libraries; e-book collaborations among university presses and nonprofit organizations; and editorial collaborations such as those recently funded by the Mellon Foundation are critically important, and among the most promising developments in the challenging and ever-changing scholarly publishing community.” In their view, innovative digital scholarly publishing, which could transform static

print or digital monographs into “vibrant hubs for discussion and engagement,” will rely on collaborative publishing models and on the adoption of sustainable open access models.

2.1.2. Systems of evaluation

These seismic shifts in scholarly publishing have engendered, many argue, a crisis for the processes of peer review and scholarly evaluation generally, because it is those processes that entrench publishing conventions (Harley et al., 2010b; Kling, 2005; Alonso et al., 2003). Fitzpatrick (2011, 2015) and others have called on the academy to recognize and adopt new systems of evaluation and authority enabled by new technologies:

Imposing traditional methods of peer review on digital publishing might help a transition to such publishing in the short term...but it will hobble us in the long term, as we employ outdated methods in a public space that operates under radically different systems of authorization.” –Fitzpatrick, 2011

Whether peer review and scholarly evaluation processes are in crisis, they have certainly lagged behind and even hindered new forms of digital scholarship (Harley et al., 2010b). A decade ago, Bates et al. (2006) deemed it essential for research in the humanities that “standards and guidelines be drawn up which will place digital resources on a sound footing and secure due recognition for the scholarly work that goes into their creation.” Despite their recommendations,¹³ and the recommendations of others (e.g., Warwick, 2007), the gap persists. The *Journal of Digital Humanities* dedicated an issue to “Closing the evaluation gap” in 2012 (Cohen and Fragaszy Troyano, 2012), which highlights among other issues the complexities of evaluating increasingly collaborative digital projects (Nowviskie, 2012) and offers potential criteria for tenure and promotion evaluation of digital scholarship (Presner, 2012; Rockwell, 2012). Mandell (2012) notes that there are venues for the peer review of innovative digital publications, including Nineteenth Century Scholarship Online (NINES) for the review of scholarship about the 19th century,¹⁴ and similar, newer organizations, each oriented toward a different era of historical and literary study. However, as Mandell notes, even the solid content and technical review provided by such organizations are not guaranteed to be recognized or accredited by other components or other agents of the processes of scholarly evaluation, e.g., tenure and promotion committees.

¹³ Study of attitudes and options for evaluation was thorough; however, their findings were specific in some respects to the context of the UK’s Arts and Humanities Research Council as a funding body.

¹⁴ <http://www.nines.org/>

[D]igital monographs without print equivalents, digital scholarship that can exist only online, or digital collections or libraries, have not received the same level of academic acceptance, either in the form of adoption by authors or recognition by peers. ... few models and tools exist for successful, sustainable, and stable all-digital publications. Importantly, those that do exist are not recognized with reviews. –McKay, 2014

The persistence of the described evaluation gap may be attributed to a complex of social factors and institutional dependencies. One acknowledged factor is a common lack of knowledge about or understanding of new genres, and how to interpret and evaluate their contributions.

Thomas (2015) urges us toward more earnest consideration of new genres of scholarship:

genres that can be circulated, reviewed, and critiqued would afford colleagues in the disciplines ways to recognize and validate this scholarship...In the next phase of the digital humanities, then, scholars have the opportunity to debate, and perhaps clarify, the qualities and characteristics of digital scholarship.

2.1.3. Library collections and collection description

In relation to questions of evaluation, thematic research collections, among other new forms of digital scholarship, exist largely outside of standard systems for the dissemination and preservation of scholarship. Thematic research collections – among other alternative forms of publication in the humanities – remain largely absent from library collections and digital repositories (Clement et al., 2013). They are not readily found in common scholarly discovery systems (such as Google Scholar, academic databases, indexing and abstracting services, etc.), and they are not usually made discoverable through libraries. The Bates et al. (2006) study cited above not only revealed widespread concern for new review and evaluation strategies; they also found that scholars were anxious about the sustainability and the legacy of their digital publications, and their exclusion from libraries:

The issue of sustainability is perceived to be of vital importance for digital resource provision. One focus group, for example, noted ‘major anxieties about the sustainability of digital resource’...

Other measures of esteem were frequently raised during the project. Inclusion in library catalogues, for example, was seen as desirable (and in a sense a form of review, since it conveys that an authoritative body considers the resource of value). –Bates et al., 2006

There is an extensive literature on adapting library collection development policies and cataloging practices to the growth in electronic books and journals, but less attention to other

digital scholarly products. Calhoun (2011) asserts the need to revitalize the library catalog, in part by connecting it to web-scale discovery tools and digital repositories, in order to enhance the visibility and relevance of library research collections. However, it is not clear what the implications are for thematic research collections and other alternative genres of scholarly publication. Horava (2011) notes that reformulating library practices of selection, acquisition, and dissemination is pivotal for academic libraries now: “Coping with the profusion of forms of scholarly publishing, variable notions of authorship, and challenges of selecting materials—all while managing a library collection budget—is no simple matter” (Horava, 2011). Horava cites a need to refocus “scoping criteria” of collection development policies – to expand interpretations of the longstanding principles of selection, including authority, originality, impact, timeliness, breadth and depths of coverage – but is less specific about strategies for coping with altogether novel forms.

Most of the literature on library practice surrounding digital scholarship addresses how libraries may take more active roles in scholarly communication, particularly prior to publication, even acting as collaborators in digital humanities endeavors (Clement et al., 2013; Jewell, 2009; McFall, 2015; Fortier and James, 2015; Caprio 2015). This is witnessed by the proliferation of library publishing programs and digital scholarship services, as discussed above, but does not engage the question of what libraries are doing or could do with new forms of digital scholarship after publication. Brantley et al. (2015) note opportunities for libraries to become involved with trends in scholarly communication, such as increasing born-digital and creative works, but focus on the benefits of institutional repositories and enhanced roles for library/faculty liaisons, rather than systemic integration of new forms of scholarship. Caprio (2015) also emphasizes the library’s role in “knowledge creating activities,” including publishing, and in new cyberinfrastructures, but does not suggest how to increase the discoverability and sustainability of materials in alternative formats or disseminated through alternative channels. Digital repositories, and institutional repositories in particular, are commonly envisioned for providing long-term access to digital scholarship (Brantley et al., 2015; Fortier and James, 2015); yet those repositories are not always integrated with existing systems of library description, representation, and discovery tools, and many cater exclusively to digital resources in conventional forms, such as articles.

A related discourse is developing on how libraries may “collect” – by cataloging – open access materials that are not created, owned, held, or licensed by the library. Several libraries have

created open-access collection development policies,¹⁵ and some systems, such as the University of California's Shared Print Program,¹⁶ routinely catalog open access journals and books that are indexed in shared directories. Emergent policies and practices in the library provision of third-party open access materials may offer models for library "collection" of thematic research collections.

The integration of thematic research collections and other new genres of digital scholarship into library collections and discovery services will rely on structured descriptions. What demands these new genres of digital scholarship will place on existing description standards is an open question, one that Tennis (2011) raises in the related context of bibliography:

Dissemination of thought in recorded form has changed. Knowledge organization, access systems, and preservation institutions have also changed, even if we focus only on their management of writings, and not other forms of recorded knowledge. Thus, if we take a broad definition of bibliography to be *the systematic enumeration and description of writings* the question surfaces, what can hundreds of years of thinking and practice of bibliography tell us about the current state of the art? Is there now a new bibliography? –Tennis, 2011

There may be aspects of innovative forms of scholarship they struggle to accommodate. There are standards for the description of collections in general, of which the most prominent is probably the Dublin Core Collections Application Profile (DC-CAP).¹⁷ The DC-CAP provides a set of terms designed to facilitate simple collection description, "suitable to a broad range of collections"; as such it is "not intended to describe every possible characteristic of every type of collection."¹⁸ Table 1 gives the DC-CAP properties.

The CIDOC Conceptual Reference Model¹⁹ (CIDOC CRM) also makes provisions for collections. Lourdi et al. (2009) offers guidance on modeling cultural heritage collections using CIDOC-CRM, and gives a mapping from DC-CAP (Kakali et al., 2007; Lourdi et al., 2009). However, by the ontological account of the CIDOC CRM (and reflecting that standard's orientation to cultural heritage institutions), collections may only be "physical objects."²⁰ Existing

¹⁵ See for example Emory's Open Access Collection Development Policy:
http://guides.main.library.emory.edu/ld.php?content_id=16498194

¹⁶ <http://www.cdlib.org/services/collections/sharedprint/>

¹⁷ <http://dublincore.org/groups/collections/collection-application-profile/>

¹⁸ <http://dublincore.org/groups/collections/collection-application-profile/#colproperties>

¹⁹ <http://www.cidoc-crm.org/>

²⁰ http://www.cidoc-crm.org/cidoc_graphical_representation_v_5_1/collection.html

schemas for the description of collections may stem from bibliographic traditions, but they were designed with cultural heritage collections in mind. It is unclear whether the same standards will suffice to describe innovative scholarly products that assume the logic of collections.

Table 1. List of DC-CAP Properties

Type	Access Rights	Date Items Created
Collection Identifier	Accrual Method	Collector
Title	Accrual Periodicity	Owner
Alternative Title	Accrual Policy	Is Located At
Description	Custodial History	Is Access Via
Size	Audience	Sub-Collection
Language	Subject	Super-Collection
Item Type	Spatial Coverage	Catalog or Index
Item Format	Temporal Coverage	Associated Collection
Rights	Date Collection Accumulated	Associated Publication

While this section has focused on libraries as the institutions primarily charged with the stewardship of digital scholarship, long-term strategies for thematic research collections and other challenging varieties of digital scholarship may rely on collaborations among alternative institutions, including independent, domain-specific data archives. Clement (2013) reaffirms the need for improved curation and preservation of digital humanities projects, including thematic research collections, but sees the solution to sustainability and preservation in a network or “collaboratory” of web-based data archives, rather than in the library exclusively.

2.2. COLLECTIONS GENERALLY

This section considers collections in general, as they are related to thematic research collections, not least by being a fundamental part of their definition.

Conceptual treatments of collections center on their ontological characterization of collections (Wickett et al., 2010; Wickett, 2012): can they be defined in terms of a familiar ontological construct, such as a set; and if not, what are they? Empirical work also studies the

concept of collection (e.g., Lee, 2000; and Roberts, 2014), along with collection uses, development, representation, and other aspects in numerous contexts: in scholarly activities, in digital libraries and aggregations, and in the context of libraries in general, where “collection” is variously used to refer to special collections, museum exhibits, archival collections, and to the whole holdings or contents of a library.

This section begins by reviewing the literature on the functions and use of collections in research, before turning to a study of conceptual approaches to collections generally that may help inform our understanding of thematic research collections.

2.2.1. Uses and functions of collections

Extensive work has been done on the development, representation, and description of research collections to support scholarship (e.g., Council on Library and Information Resources, 2010; Hill et al., 1999; Palmer et al., 2010; Wickett et al., 2013; Sinn & Soares, 2014). Research on collection representation and structure in digital libraries and on the Web suggests a number of functions: to support navigation (Lee, 2000; Lagoze and Fielding, 1998), to provide context for items (Wickett et al., 2013), and to improve subject access to items (Zavalina, 2010).²¹ Empirical evidence of collection use reaffirms the navigational functions of collections. Johnston and Robertson (2002) show that “the existence of collection-level descriptions supports the high-level navigation of a large (and perhaps distributed and heterogeneous) resource base.” Humanities researchers, in particular, have demonstrated reliance on collections as research platforms, where “platform” often entails some navigational function (Palmer, 2004; Palmer, 2005; Brogan, 2006; Dempsey, 2006; Mueller, 2010; Green and Courtney, 2014). Several studies show that institutionally curated collections are most useful at the outset of humanities research, suggesting that collections are particularly useful for navigating the information universe, finding, and selecting relevant materials (Duff & Johnson, 2002; Tibbo, 2003; Buchanan et al., 2005; Palmer, 2004; Palmer et al., 2009). Assuming collections are topically coherent, they also provide a strong

²¹ In addition, there is some evidence that collection information can improve topic modeling of aggregate metadata records. The subject analysis team of the IMLS Digital Collections and Content project found that language models for collections could be exploited to improve estimates of language models for items, in the process of topic modeling across the aggregate. They hypothesized that the very fact that documents are selected and gathered into collections is informative (Efron et al., 2011).

browsing layer. Zavalina (2010) shows that collection-level subject access is a powerful mechanism in large-scale digital libraries. Even browsing the results of a search, in a large-scale digital library, can be overwhelming if items are decontextualized. Wickett et al. (2013) demonstrate how adding a collection-level browsing layer to search results provides a more intuitive view of the topical landscape of large-scale digital libraries. It is unclear whether all of these functions of library and digital collections are replicated by scholar-created collections.

Collections are fundamental to the activities and processes of humanities research. Humanities scholars are known to gather information from various sources as an essential, often preliminary step in the research process (Palmer & Neumann, 2002; Palmer, 2004; Sukovic, 2008; Sukovic, 2011; Toms & Flora, 2005; Toms and O'Brien, 2008). They “build their own personal libraries to support not only particular projects but also general reading in their field,” largely out of a need for constant, convenient access to materials for rereading or analysis (Brockman et al., 2001; Palmer, 2005). Palmer et al. (2009) identify “gathering” and “organizing” as primitives of the scholarly “collecting” activity. Scholars’ personal research collections include both primary and secondary sources, in numerous media and formats drawn from heterogeneous sources (Brockman et al., 2001; Palmer & Neumann, 2002; Palmer, 2005). Mueller (2010) employs the metaphor of the library carrel to describe how digital humanities scholars collect texts and subsets of texts that are amenable to computational analysis. Indeed, a survey of scholars working with large-scale text corpora found that they want improved ways of finding and handling relevant subsets of the corpora:

Researchers do not necessarily need huge sets of data to do interesting work, but the implication is that they do need flexible data delivery services that can deliver different kinds of data in different formats based on different searches for different kinds of research at different times. —Varvel and Thomer, 2011

User-generated collections more generally have been treated in studies of how users retrieve and synthesize materials from digital libraries (Feng et al., 2004); personal data collections (Beagrie, 2005); preservation of faculty-created digital collections (Beaudoin, 2011); collections of photographs on Flickr (Stvilia et al., 2009; Rorissa, 2010); and in one study of journalistic research practice (Attfield & Dowell, 2003). Beagrie (2005) does note the high potential value of scholarly collections: “their importance for current scholarship is growing along with the power and reach of software tools and communications available to individuals to create, manage, and disseminate them.”

2.2.2. Conceptions of collections

Svenonius lists collections among the fundamental bibliographic entities, defining “collection” as, “a set of documents gathered on a basis of one or several attributes to be described collectively” (Svenonius, 2000). The description of collections as sets, whether casually or formally, is common but contested; indeed, there seem to be no widely accepted conclusions about the ontological status of collections (Wickett et al., 2011). One account of collections, as *sets in a curatorial role*, suggests *curatorial intent* – or the intention or attention of a person or agent in a curatorial role – as a condition for collection-hood (Wickett et al., 2011; Renear et al., 2008; Wickett, 2012). Indeed, selection (which may be understood as a manifestation of curatorial intent) is an implicit or explicit feature of various conceptual accounts of collections (Lagoze and Fielding, 1998; Lee, 2000; Flanders, 2014). This intuitive feature is necessary to any account of thematic research collections, which are defined as products of scholarly (curatorial) work. Indeed, if thematic research collections are a special type of collection, they may place an even higher demand on curatorial intent: not only that it exist, but that it be of a particular sort – an intention *to support research on a theme*.

Despite the ontological ambiguity of the collection, Corral and Roberts point to a high degree of shared understanding of the concept by library professionals, users, and even non-users of libraries (Roberts, 2014; Corral and Roberts, 2014). Their empirical study identifies three prevalent concepts of the library collection, each with its own implications for collection development: collection as thing (e.g., a group of materials), which is the most common understanding; collection as access (collection as connection); and collection as process (e.g., as selection, as search, as service). While alternate conceptualizations may expand how libraries conceive and develop collections, the question remains unaddressed of whether these conceptions might translate to a more general understanding of collections (absent institutional context) or the more specific genre of thematic research collection. Indeed, most of the conceptual literature on collections pertains to institutionally developed collections (Lee, 2000, 2005; Johnston and Robinson, 2002; Corral and Roberts, 2012).²² While a comparison of those conceptions with

²² In case the distinction between institutional and scholar-generated collections is not intuitive: “There is also a worthwhile distinction to be made between resources produced within academia, and those created by bodies in the museums, libraries and archives sector. Such resources will have been developed under different imperatives, with a focus primarily on knowledge transfer rather than on research. While it is clear that many resources in this category involve significant academic input, and quality-assurance mechanisms such as steering and user groups will be

conceptions of scholar-generated collections may prove fruitful, the latter have roots in research processes that provide richer context for the concept.

Despite the resemblance between digital collections and collections in the historical sense (rooted in physical collocations), Flanders (2014) acknowledges the “distinctive epistemological conditions under which they present themselves to us.” Of course, many digital collections originated as representations of physical collections, and thus the genre as a whole may inherit features and limitations of physical collections (Flanders, 2014). However, digital collections characterize a shift to a new “digital research ecology,” which is oriented toward aggregation. In this new ecology or infrastructure, individual items must be understood as contextualized by metadata and by search and navigational functions at the collection level, “mechanisms that do not arise as part of the rhetoric of the individual text but rather are constituted as informational layers that may operate independently of any single text” (Flanders, 2014). Adapting terms from Ramsay (2014), Flanders invites us to reconceive the digital collection, not as a *network* of preexisting and commensurable information resources, but as a crafted, *patchwork* assemblage, in which collection *actively* serves to relate previously unrelated and incommensurable items. This view highlights the digital collection as a venue for scholarly discourse, distinctive in purpose and form from a library collection:

If the patchwork collection thus acknowledges its manufactured quality, then it can also help us understand the collection as both expressing and supporting analysis ... the patchwork collection supports analysis, through that same explicitness and transparency, by permitting a distinctively important kind of intellectual transaction: not the all-sufficiency of traditional scholarly product that seeks to say everything itself, and not the passivity of the library that seeks only to ‘support’ and be raw’, but a give and take, a negotiation of meaning that reminds us that scholarly inquiry is always a transaction involving agency on both ends” -Flanders 2014

2.3. THEMATIC RESEARCH COLLECTIONS

The thematic research collection has long been acknowledged as a genre of scholarly production in the humanities (Unsworth, 2000; Brockman et al., 2001; Alonso et al., 2003; Palmer, 2004; Schreibman et al., 2008; Ciula & Lopez, 2009; Price, 2009; Flanders, 2014; Meiman, 2015; Thomas, 2016). In 2004, Palmer predicted their rise: “scholar-created research collections are

integral to their development, they are qualitatively different from resources funded by the UK research councils, and are not generally subject to the same type of initial, formal peer review” (Bates et al., 2006).

likely to increase in number as the work of producing them becomes more widely accepted as legitimate scholarship” (Palmer, 2004). However, the literature on thematic research collections as a form of alternative scholarly publishing remains sparse, despite their rising number and increasing demands that this and other digital genres be valorized in scholarly evaluation processes (Harley et al., 2010b; Rockwell, 2011; Modern Language Association, 2012; Fenlon et al., 2014; AHA, 2015).

2.3.1. Characteristics of thematic research collections

In the most thorough account of thematic research collections, Palmer (2004) develops and expands upon Unsworth’s (2000) list of endemic features: they are digital, thematically coherent, heterogeneous, structured, open-ended, designed to support research, and authored or multi-authored. They function to support research, and beyond that, they represent a scholarly contribution, (Palmer, 2004). Some thematic research collections aim to serve as platforms for interdisciplinary research, and some offer tools to support research activities (Palmer, 2004). In addition, thematic research collections are hypothesized to exhibit contextual mass (Palmer, 2004; Palmer et al., 2010; Clement et al., 2013; Green and Courtney, 2014; Flanders, 2014).

Contextual mass is a posited development principle for digital collections, libraries, and aggregations. A collection with contextual mass is one in which items have been purposefully selected, organized, and bestowed with sufficient context to support deep, multifaceted inquiry on a theme (Palmer et al., 2010). The concept is an intuitively appealing one; Green and Courtney (2014) argue that contextual mass “is more imperative than ever in the development of digital library collections,” as it reflects an active user-orientation to development. “Contextual mass” has not been precisely defined, but some dimensions of collections have been associated with the concept: density, cohesiveness, interconnectedness, and diversity or heterogeneity (Palmer et al., 2010). Palmer et al. (2010) found a number of ways of measuring or operationalizing these dimensions, within the context of a massive aggregation of cultural heritage metadata, in order to discover subject specializations or themes within the aggregation that obtained contextual mass. However, not all of their measures – such as the number of small vs. large collections represented within a subject specialization – are applicable outside of the context of digital library aggregation. Taking a step back, however, we can see a pattern in their analysis: that of cohesiveness or thematic density, offset by heterogeneity or diversity of evidence. It could be a balance between these

contrasting factors that characterizes a collection of contextual mass, or a rich collection of humanities data.

Palmer (2005) explores thematic research collections as a new kind of access resource, or tertiary resource for the discovery and evaluation of publications and other information sources. Seeing thematic research collections as scholar-created access resources, in the vein of bibliographies or literature reviews, highlights their duality as both scholarly products and platforms for new research:

However, scholars are not only constructing environments where research materials can be accessed more conveniently by more people, they are also performing their normal scholarly role of creating research products that advance the state of scholarship in the field. Like other scholarship in the humanities, research takes place in the creation of the work, and research is advanced because of it. –Palmer, 2005

Thomas (2016) identifies thematic research collections as “perhaps the most well-defined genre in digital humanities scholarship,” characterizing them as “sprawling investigations” that bring together archival materials and tools, and embed “interpretive affordances” into a collection. Thomas situates the thematic research collection among two other perceived genres of digital scholarship in the humanities, the interactive scholarly work and the digital narrative, differentiating the genres as shown in Figure 3 (Thomas, 2016). By Thomas’ account, thematic research collections are differentiated from interactive and narrative works by being capacious in scope, as opposed to tightly defined or problem-oriented. Existing characterizations of thematic research collections make no claims about the size of the collection or scope of its theme, though “capacious” may be an apt way of describing the genre’s duality as both scholarly product and platform for scholarship, and its balance of thematic coherence with contextual mass. In addition, Thomas suggests that thematic research collections offer affordances for interpretation rather than being explicitly interpretive, though he suggests that the “next phase of thematic research collections might feature interpretive scholarship embedded within and in relationship to the collection” (Thomas, 2016). Positioning thematic research collections in the history digital humanities scholarship over the past 20 years, Thomas calls for further clarification of the genre:

In this first phase of the digital humanities, scholars produced innovative and sophisticated hybrid works of scholarship...Although such experimentation should continue, genres that can be circulated, reviewed, and critiqued would afford colleagues in the disciplines ways to recognize and validate this scholarship.

Properly focused but broadly defined, such genres might alter the disciplinary conversation and appear in venues that provide a foundation for future scholarship in the disciplines.

	Interactive scholarly works	Digital projects or thematic research collections	Digital narratives
Type of data	Homogeneous, primary	Heterogeneous, primary	Integrated, layered
Components	APIs, scripting	Schema, data models	Analysis, modules
Organization	Hypothesis	Theme or subject	Criticism
Scope	Tightly defined	Capacious	Problem-oriented
Interpretative nature	Query-based	Affordances	Multimodal
Character	Procedural inquiry	Open-ended	User-directed, hypertextual
Examples	ORBIS (orbis.stanford.edu) <i>Visualizing Emancipation</i> (dsl.richmond.edu/emancipation) <i>Railroaded</i> (railroaded.stanford.edu) <i>Who Killed William Robinson?</i> (canadianmysteries.ca/en/robinson.html) <i>O Say Can You See: Early Washington DC Law and Family</i> (earlywashingtondc.org)	<i>Valley of the Shadow</i> (valley.lib.virginia.edu) <i>Whitman Archive</i> (whitmanarchive.org) <i>Mapping the Republic of Letters</i> (republicofletters.stanford.edu) <i>Digital Gazetteer of the Song Dynasty</i> (songgis.ucmerced.edu)	<i>The Differences Slavery Made</i> (www2.vcdh.virginia.edu/AHR) <i>Gilded Age Plains City</i> (gildedage.unl.edu) <i>Who Shot Liberty Valance?</i> (mamber.filmvtv.ucla.edu/LibertyValance) <i>Hearing the Music of the Hemispheres</i> (scalar.usc.edu/anvc/music-of-the-hemispheres) <i>Queering Slavery Working Group Tumblr</i> (http://qswg.tumblr.com)

Figure 3. Thomas' (2016) matrix of digital humanities scholarship

2.3.2. Primary sources and humanities data curation

One aspect that appears to distinguish thematic research collections from more familiar genres of scholarly production is the priority they place on the primary source (Palmer, 2004). Studies have shown the increasing prevalence of digital primary sources and their changing use and presentation in scholarship (Brockman, 2001; Palmer, 2005; Green and Courtney, 2014; Schöch, 2013). Unlike the monograph or journal article – which may include reproductions of the primary sources that serve as their evidence, but which foreground narrative interpretation of the evidence – thematic research collections foreground the evidence itself. They function to make primary sources and their contexts highly visible (Palmer, 2010), and while they may attend

sources with narration, argumentation, or explicit interpretation, much of the scholarly work of thematic research collection inheres in the selection and representation of sources.

The American Council of Learned Societies (2006) report on cyberinfrastructure for the humanities and social sciences asserts that “[d]igital cultural heritage resources are a fundamental dataset for the humanities.” They describe digital collection-building as central to the future of digital scholarship in the humanities. If we consider primary sources, such as cultural heritage resources and texts, to be humanities data, it is worth considering thematic research collections as participating in and as subject to data curation in the humanities.

According to Flanders and Muñoz (2012), the term *curation* “carries this dual emphasis: on protection, but also on amelioration, contextualization, and effective exposure to an appropriate set of users.” Thematic research collections manifest curatorial intent, as we have described, but this sense of “curatorial” leans toward the latter aspects, of contextualization and exposure. Despite bearing designations as “archives” or “repositories,” thematic research collections do not in general prioritize the preservation or stewardship of sources over the long term. Nonetheless, aspects of curation are borne out in their development: in the selection of sources as relevant to a theme and worthy of scholarly consideration, and in the organization, contextualization, and presentation of those sources (Palmer, 2004; Mandell, 2012). Putting this in terms of the definition of data curation, most thematic research collections do not take responsibility for the “active and on-going management of data through its lifecycle,” but their development can be described “activities [which] enable data discovery and retrieval, maintain quality, add value, and provide for re-use over time” (Cragin et al. 2007). These aspects of thematic research collections may yield insights beneficial to the praxis of data curation in the humanities, as well as to the development of other kinds of collections, as Palmer (2004, 2005) and Green and Courtney (2014) have noted. There is work to be done in understanding the intersection of collection with curation, as Flanders & Muñoz (2012) suggest.

In turn, thematic research collections themselves are subject to curation, in their role as scholarly products. Many thematic research collections readily meet Schöch’s (2013) definition of humanities data, as “a digital, selectively constructed, machine-actionable abstraction representing some aspects of a given object of humanistic inquiry.” Flanders and Muñoz (2012) raise this duality of thematic research collections, as both curating humanities evidence (if in a limited sense), and as being products desirous of curation:

...humanities data is presented in specialized aggregations that themselves have significance for understanding, using, and curating the data. Some of these aggregations are digital extensions of long-standing traditional forms: for instance, finding aids, concordances, and scholarly editions, which have a long analog history. Others, like the thematic research collection or digital text corpus, are products of new digital research methods... –Flanders and Muñoz, 2012

As such, thematic research collections entail unique requirements for digital curation. They bind text data, images, and contextual information together in “highly structured ways”; while these collections are aggregations of the organization and “editorial logic that is represented in ancillary materials such as stylesheets and configuration files is likely to be extremely significant,” both for sense-making of the collection and for recovering the curatorial intentions that “constitute it as scholarship” (Flanders and Muñoz, 2012).

Green and Courtney (2014) find that there is a growing sense among humanities scholars that humanities “datasets” – whatever shape they may take – constitute publishable scholarly contributions. What relationship thematic research collections bear to humanities data sets is worth further exploration. Muñoz (2013) makes the conceptual link between publishing and data curation, which nexus thematic research collections occupy. Publishing humanities data and linking humanities publications to relevant datasets are central goals of another emergent genre of digital scholarship: enhanced publications.

2.3.3. Enhanced publications and research objects

Enhanced publications are publications of scholarly narratives enriched with embedded or linked supplementary content, such as data sets, multimedia materials, related resources, facilities for annotation or commenting, and opportunities for interactive or alternative modes of presentation or reading (Woutersen-Windhouver and Brandsma, 2009; Jankowski, 2012; Bardi and Manghi, 2014). Research and development of enhanced publications builds on the extensive literature on advancing scholarly communication across disciplines in the advent of data-intensive scholarship and related, enabling technologies, not least the emergence of linked data and semantic metadata standards (e.g., Van de Sompel, et al., 2004; Bourne, et al., 2011; Bechhofer et al., 2013; Assante, et al., 2015). Enhanced publications aim to contextualize scholarly narratives with persistent, meaningful connections between data sets, research processes, and associated resources and publications, and at the same time enable validation and reproducibility of scientific and computational results. In fields not oriented toward data-centric or reproducible research, goals include enabling scholars to share more diverse media, to convey their interpretations and

arguments in more complex and representative ways, and to semantically interrelate sources and references with narratives. Jankowski, et al. (2012) found several other motives of authors engaged with authoring enhanced publications, including creating dynamic spaces for ongoing collaborative authorship, creating community around publications, serving further research processes, and promoting publications.

Sigarchian et al. (2014) relate a set of functional goals of enhanced publications to a set of desired attributes drawn from a survey of the literature (see Figure 4), with the objective of comparing the utility of different data models for representing enhanced publications. For our purposes, their organization of attributes and functional goals offers a concise summary of the range of features that may be present in the genre. The more an enhanced publication includes or perhaps even foregrounds primary sources and related content over its narrative base, the more it begins to resemble what we have conceived as a thematic research collection, especially in light of the research and collaboration objectives of enhanced publishing.

The literature has not yet explicitly related enhanced publications to thematic research collections as such, though Breure et al. (2011) locate specific resources, which are recognizable as examples of thematic research collections according to our definition, on a proposed spectrum of publication types. This spectrum, from “conventional” to “rich internet applications,” is arranged according to the amount and quality of enhancements made to the publication, such as interactive and multimedia elements, and non-linear modes of reading and exploration (see Figure 5). Breure et al. categorize things in the vein of thematic research collections as “Type II Rich Internet Publications,” which may be more recognizable as “interactive multimedia applications” or “experiments in digital scholarship” than as publications in any conventional sense.

Functional Goals	Attributes	Format			
		PDF	HTML	EPUB2	EPUB3
Packaging with supplementary material	<ul style="list-style-type: none"> – Embedding and linking of multimedia/interactive objects – Document integrity and structure – Attached data resources – Navigating to an embedded / linked media object 	✓*			✓
Enabling repetition of experiments	<ul style="list-style-type: none"> – Native support for interactivity – Code execution – Interactive content 		✓		✓
Improving readability and understanding	<ul style="list-style-type: none"> – Navigation by table of contents – Reader comments – Appearance – Page transitions – In-page navigation – Image browsing – Links to figures and tables – Direct publication links from references – Cited by 	✓*	✓*	✓	✓
Interlinking with research data	<ul style="list-style-type: none"> – Metadata – Relations 		✓	✓*	✓
Portable packaged file	<ul style="list-style-type: none"> – Bridging online / offline – Transmission 	✓*		✓*	✓
Suitable for print	<ul style="list-style-type: none"> – Download as PDF 	✓			

Figure 4. Support for enhanced publication attributes and functional goals (*=limited support) (Sigarchian et al., 2014)

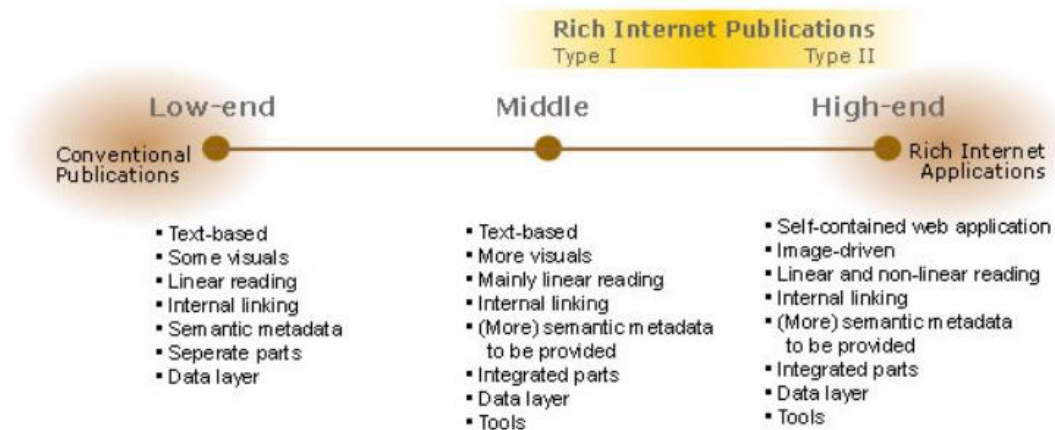


Figure 5. Kinds of enhanced publication (adapted from Breure et. al, 2011)

Enhanced publications, and especially those that may be seen to fall into the type of *rich internet publication*, share fundamental, perhaps even definitive qualities with thematic research collections. Both genres aggregate and meaningfully interrelate heterogeneous components. The data models that support their representations are the same or similar (as I will discuss below). And both genres confront significant, systemic challenges, such as the difficulty of ensuring their discoverability, due to inadequate descriptive standards, and the difficulty of ensuring long-term maintenance of complex, compound digital objects (Woutersen-Windhower and Brandsma, 2009; Bardi and Manghi, 2015).

However, enhanced publications bear some important distinctions from thematic research collections. First, accounts of enhanced publications deemphasize the curatorial aspect of production, of the selection and gathering of sources that serve to enrich scholarly narratives. They are not considered to be collections, despite their resemblance, because they are still, by definition, grounded in scholarly narratives. Yet, as this dissertation will show, the boundary between narrative and collection can be fuzzy; many thematic research collections employ narrative as an interpretive layer on top of a base of sources. Second, thematic research collections have been in development for a couple of decades now and there are some established patterns of funding and collaboration to produce collections. In contrast, production of enhanced publications is comparatively less widespread. There have been several proof-of-concept projects in the sciences and humanities, including infrastructure-building projects; and many publishers have experimented with or fully adopted certain enhancements to their otherwise conventional, digital publications. However, from a systemic perspective, it remains unclear where the burden of development of enhanced publications should fall. For example, Breure et al., 2011, question the extent to which publishers as opposed to authors should assume responsibility for enhancement. In other words, enhancement seems to be perceived as additive rather than inherent to the process of production of this genre of scholarship, with the value of certain kinds of enhancement remaining questionable for some authors (Jankowski, 2011). Finally, the most fundamental difference is that these products appear to be motivated by different reasons, at least on the surface: enhanced publications to publish (finished) narrative scholarship, and thematic research collections to support research on a theme.

A second genre of production that is essentially similar both to enhanced publications and thematic research collections is the *research object*, defined as a semantically rich aggregation of

resources assembled to support a research objective (Bechhofer et al., 2010; Bechhofer et al., 2013). Research objects are increasingly employed for the representation of scientific workflows, and they have begun to see application in the representation of computational workflows and research objects in the digital humanities (Almas, 2017; Page, Lewis and Weigl, 2017). What is the difference between a curated collection, designed to support research on a theme, and a “principled aggregation of resources,” which possesses “scientific intent” (Bechhofer et al., 2010)? The differences may be more contextual than conceptual. In addition, a distinctive goal of research objects is to make objects and workflows machine-actionable; this is not an ostensible goal of thematic research collections so far, but it is not an inconceivable prospect, particularly in the context of computational digital humanities work.

Despite their differences, it is worth exploring the significant areas of overlap among these three genres. Chapter 6 describes the potential implications of research objects and enhanced publication data models and management systems for the sustainability and preservation of thematic research collections (see section 6.4). Chapter 7 describes future work on how this study of thematic research collections may refine enhanced publication data models for the representation and management of scholarly collections.

CHAPTER 3: METHODS

3.1. OVERVIEW OF APPROACHES

To recapitulate, my research questions are:

- (R1) What are the defining features of thematic research collections as a scholarly genre?
- (R2) What are the challenges, for libraries and related scholarly-publishing entities, in supporting thematic research collections as a scholarly genre?

I approached (R1) with a provisional typology of thematic research collections, supplemented with a content analysis of selected exemplars of resulting types of collections. Sections 3.2 and 3.3 detail these two methods, respectively. I approached (R2) using a set of interviews with representatives of digital humanities centers and libraries. Section 3.4 gives details for this approach.

A provisional typology of a large sample of collections afforded a broad view on the landscape of thematic research collections. Distinguishing collections by their underlying data models, the typology suggested five provisional types. Exemplars of those types were selected from the broad sample of collections, and subjected to qualitative content analysis. The content analysis gave a deeper view of each provisional type of collection and how they were distinguished not only by data models, but by overarching purposes.

Content analysis revealed how collections are shaped by their different purposes. I used results of the content analysis to refine the typological analysis, resulting in a final typology of three types of thematic research collection. Together, the content analysis and typological analysis afforded some insight onto what sets thematic research collections apart as a genre: what attributes help to define collections, distinguish them from one another, and determine their contributions to scholarship.

Finally, I conducted a set of interviews with representatives of digital humanities centers and libraries to shed light on challenges to supporting the genre, strategies for addressing those challenges, and roles that institutions and individuals play in these strategies.

Figure 6 summarizes how these methods shed light on my research questions, and points to the relevant chapters of this dissertation. Results of the provisional typology informed initial protocol development for the content analysis, and the selection of exemplars. Content analysis identified several purposes of collections, detailed in Chapter 4, and then fed back into the

typological analysis in order to produce the final set of types, discussed in Chapter 5. The dashed and dotted lines in Figure 6 represent less direct but still important contributions of each method to other aspects of the study: interview data expanded and contextualized my sense of collection purposes and kinds, Chapter 5, and the outcomes of the typology and content analysis helped to clarify and exemplify sustainability and preservation challenges described in Chapter 6. In addition, the initial survey of collections conducted for the provisional typology informed the sample of collections used for content analysis and helped expand and clarify the interview protocol.

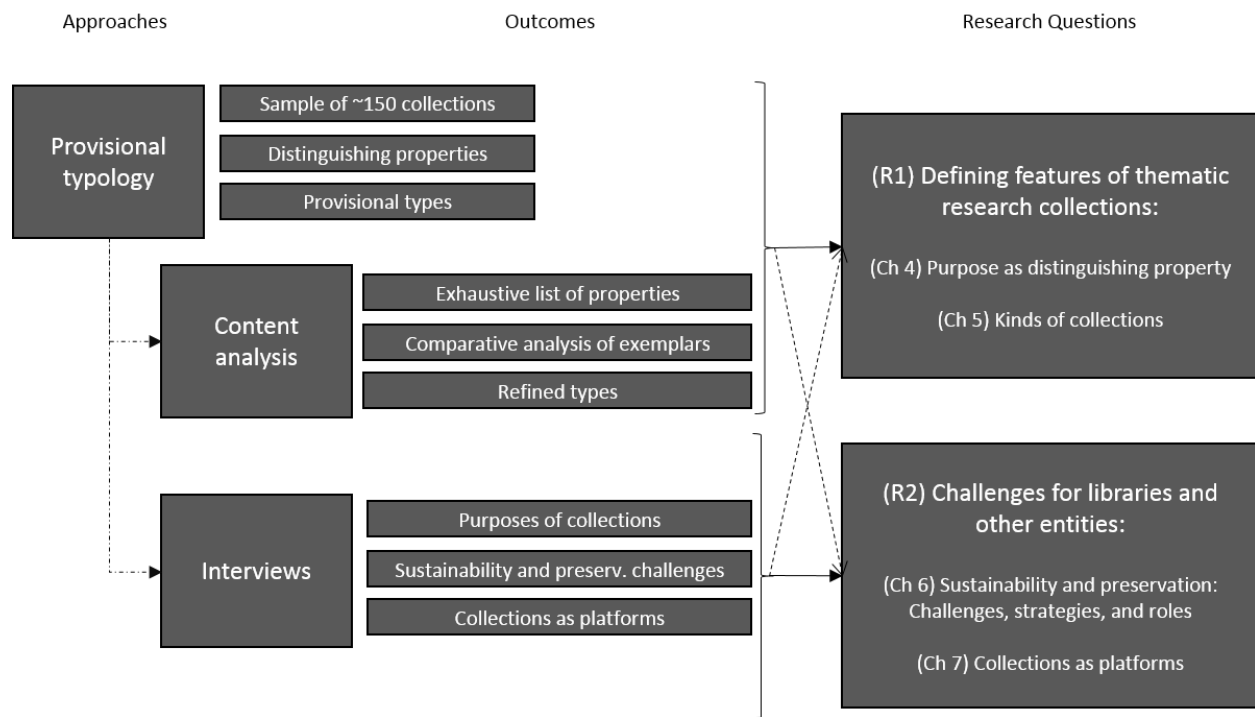


Figure 6. Approaches mapped to outcomes, research questions, and chapters herein

The rest of this section gives, for each of my approaches, a purpose, an overview (encompassing design, analysis, and sources), and limitations.

3.2. PROVISIONAL TYPOLOGY OF COLLECTIONS

3.2.1. Purpose

This typology aimed to expand my understanding of the breadth and variety of thematic research collections. The typology did not aim to define ground truth, either of what thematic research collections are, or of what sorts they are. Rather, it established an analytical framework to ground and support deeper analysis. By surveying a large sample of things that appear to meet our current characterization of thematic research collections, I gained a sense of the perimeters of the genre, the diversity of things occupying it, and how it bleeds into related genres.

There exists a wide range of things that meet our working definition of thematic research collections.²³ It seems intuitively true that the diversity might usefully resolve into kinds or types. *Typology* is a formal methodological tool for the organization of our thoughts about the reality of objects or events, a way of organizing the members of an identified class. It is a kind of classification work (Marradi, 1990), which aims to group the members of a set by some identified properties. Properties are chosen and groupings are made in such a way as to maximize both homogeneity within groups and mutual heterogeneity between groups (Marradi, 1990; Kluge, 2000).²⁴ The properties that differentiate groups of objects from one another are not essential to objects, but chosen to suit the purpose of the typology; as Koch (2000) notes, different typologies of the same class of objects might support different goals.

Kluge (2000) offers a summary of how formal typology generally proceeds,²⁵ which I adapt with relevant examples:

²³In pilot work for this study, I found that locating thematic research collections for analysis was more difficult than anticipated, not because they were rare, but because digital humanities centers (in their capacity as content-hosts or publishers) and other platforms offer so many things that meet or come very close to meeting our existing definition of “thematic research collection.” This led me to question what I should include in the study, and to typological work as an inevitable first step toward refining our understanding and definition of the genre.

²⁴This maximization is not always absolute: “we argue that the criterion of establishing mutually exclusive categories provides a useful norm in constructing typologies. Yet not all analytically interesting typologies meet this standard” (Collier, 2008).

²⁵ Similarly, by Marradi’s (1990) account, the first two things that must be established to ensure typological rigor are (1) membership of the set to be subdivided, and (2) array of properties in terms of which the internal homogeneity and mutual heterogeneity of classes are to be maximized. After this, Marradi requires a series of further establishments, including procedures for identifying properties, logical formulas for combining the differences on all properties, and decision rules on how to form classes. I have mentioned that even getting as far as Marradi’s step (1) has been difficult in pilot efforts for this project, and has represented a certain level of typological work – that involved in circumscribing the genre in the first place.

1. Identify the class of objects to be “typed” and its members. In this study, the class of objects is the class of thematic research collections; its members are individual thematic research collections.
2. Develop relevant analyzing dimensions or properties, the bases of division (Marradi, 1990; Blackburn, 2008). There is an enormous number of properties that could distinguish types within our class of thematic research collections. If we were to choose properties and groupings only to obtain mutual exclusivity of types, we could do so trivially, and with uninteresting results. The justification for choice of properties relies on common intuitions about and literature on collections and their use; they are meant to identify *interesting* differences between collections, within this context of scholarly work and use of collections.
3. Group the members by the relevant properties.
4. Analyze meaningful relationships and construct types.
5. (Repeat earlier steps as necessary to accommodate things that do not fit.)
6. Describe and name constructed types.

This study adds a final two steps to this method:

7. Closely analyze exemplars of provisional types using qualitative content analysis.
8. Repeat earlier steps as necessary to refine, describe, and name types.

A typology may proceed by the construction of a matrix or a table, as in example in Figure 7.

	Properties:		
Cases:	Property A	Property B	Property C
Case 1	✓		✓
Case 2		✓	✓
Case 3	✓		
Case 4		✓	✓
Case 5	✓		

Figure 7. Example of generic typology construction

In the example given by Figure 7, colors of rows identify groupings by unique combinations of properties. They are (Case 1), which has properties A and C; (Case 2, Case 4),

which have properties B and C but not A; and (Case 3, Case 5), which only exhibit property A. These groupings are potential *types*. Potential types are checked for whether they account for all cases, and whether they reconcile with our evolving intuitions about the features of our cases (collections). If not, new properties are identified to divide collections into more useful types. Below, in “Design Overview,” I describe how this process of iterative development went in this study.

Typologies serve a number of purposes in LIS practice and research. In information systems, informal typologies are rampant. They serve to support discovery. The most obvious examples are bibliographic classification systems and faceted browsing structures for digital libraries. In LIS research, typologies are employed to elaborate concepts. Witness abundant typologies (or analyses of typologies) of things ranging from information systems (Kakar, 2016), information retrieval systems (Ortega, 2012), and libraries (Maistrovich, 2014), to librarians (Vanwynsberghe et al., 2015), uses and users (Fleming-May, 2011), games (Pe-Than et al., 2015), documents (Pejšová and Vaska, 2011), and even information itself in different domains (e.g., Rousi et al., 2016). While typologies themselves do not make ontological assertions, they may be effective precursors to ontological work. In a discussion of a typology of online subject gateways, Koch identifies the following uses: “Typologies allow the understanding of the breadth and variety of already existing services and support their description...Typologies might be used to discover missing variations which could be worthwhile experimenting with. Typologies can help us to determine if different approaches and solutions for the various services are needed” (Koch, 2000).

These uses – understanding the breadth and variety of a genre, identifying variations missing from conventional conceptions or analyses of a genre, and discovering gaps in service to the genre – can be extended from Koch’s subject gateways to all kinds of information objects, technologies, and services. Not least, we are in need of such understandings about thematic research collections.

For the purposes of this study, the production of a complete, formal typology of thematic research collections was unnecessary. A formal definition would, for example, provide necessary and sufficient conditions for membership in a class or type (Marradi, 1990). Even assuming formal definition is possible, this study did not require that level of analysis to ground the next stages of work. The types of collections that resulted from typology (augmented with content analysis),

discussed in Chapter 5, are intended to suggest broad patterns – of how collections are built to serve different kinds of purposes for scholars – rather than strictly exclusive categories.

3.2.2. Overview

I began by identifying a sample of thematic research collections from the following sources:

- Digital humanities centers: This study examined collections from each of the centerNet Founding Centers in the United States, including the Center for Digital Research in the Humanities (University of Nebraska-Lincoln); the Center for Digital Scholarship (Brown University); Maryland Institute for Technology in the Humanities (University of Maryland); Matrix (Michigan State University); Roy Rosenzweig Center for History and New Media (George Mason University); and the Scholars' Lab (University of Virginia Library). The centerNet Founding Centers were chosen to limit the survey because centerNet is an international network of digital humanities centers, and its Founding Centers represent a prestigious and well established subset of that network. I limited the survey to U.S. centers because this study is oriented toward scholarly communication in the U.S. context. In addition to these centers, the study surveyed collections from the Institute for Advanced Technology in the Humanities (University of Virginia), because that institution was the only institution represented by an interview participant (section 3.4, below) but not included among centerNet Founding Centers.
- Tools/platforms for publishing and communication. I relied on Zorich (2008, Appendix F), which identifies tools in use by humanists, and added select tools that have been developed or which obtained relatively widespread use after the publication of this survey, including Omeka²⁶ and Scalar.²⁷
- Scholarly collective/Peer-review organizations for digital publications, including Nineteenth Century Scholarship Online (NINES).²⁸

I examined the sample to identify a set of properties that can divide collections. I started by looking, simply enough, for four properties entailed in our definition of “thematic research collection”: a collection (1) gathers primary sources; (2) demonstrates scholarly effort; (3) is

²⁶ <https://omeka.org/showcase/>

²⁷ <http://scalar.usc.edu/>

²⁸ <http://www.nines.org/>

thematic; (4) supports research on a theme. While everything in the final sample evinced these properties enough to be included in the study, I was struck by how difficult it was to determine, sometimes, whether they did. Therefore I began to focus on “edge cases,” which thwart our traditional conceptions of collections, as opposed to “traditional” collections, which were readily identifiable or self-described as thematic research collections. For example, many edge cases did not make primary sources immediately or obviously accessible through direct search and browse. What is an “item” in a collection that does not provide direct search and browse across discrete primary sources? And what is a collection without readily identifiable items? Second, the more “traditional” collections were conveniently differentiated by whether they are text-based collections that invested heavily in advanced markup of their texts. Both of these aspects – of being built around advanced markup or not, and providing direct or indirect access to primary sources – stem from the data models underlying collections.

Therefore, provisional analysis relied on properties pertaining to the data models of collections because those models serve to embody scholarly interpretation, help determine potential uses of collections, and affect their long-term accessibility and maintenance. Cases were grouped by the presence or absence of properties. The resulting groups represented preliminary types of thematic research collections, and the first outcome of this research. This outcome is discussed further in section 5. 2.

The specific purpose of provisional typology was to inform the second and third phases of work in the following ways: First, the typological analysis showed that our working definition of *thematic research collection* encompasses a diversity of digital resources. Second, the properties used to distinguish collections in the provisional typology – revolving around collections’ data models – helped lay a foundation for deeper inquiry into how the design of collections reflect and implement their intended contributions. Third, an improved sense of the full range of the genre and what it encompasses allowed for a selection of diverse representative collections for close analysis, expanded the scope of the content analysis while also highlighting the most potentially interesting features of collections.

Content analysis served to refine the properties that determined the types, giving me a richer sense of the purposes of collections and how they help shape collection development. After content analysis and the refinement of three different constellations of properties of collections (purpose, completeness, theme, items, diversity, interrelatedness), approximately 45 further

collections were subject to typological analysis, resulting in a typed sample of 145 collections total. The resulting typology, applied to the sample, is available in Appendix A, and discussed at length in Chapter 5.

3.2.3. Limitations

I have stated that this typology was purposefully limited in scope. I did not aim for formal definition of types of collections, or conclusive and complete representation of the universe of thematic research collections. This work was meant primarily to serve the other stages of this study. It seems endemic to this genre, which is experimental and genre-bending, that any attempt at definition must be qualified by exceptions.

It would have been impossible to sample from the whole population of extant thematic research collections because they are difficult to find and identify – in part because they are not always called as such, and they are not categorically discoverable through information systems like library catalogs. There does not seem to be any agreement about how to circumscribe the genre in any case.

Finally, typologies do not make ontological assertions. This is what Marradi (1990) terms the “essentialist fallacy”. We will not be able to assert, at the end of this, that any typological distinctions we produce are ‘natural’, ‘inherent’, ‘essential’, or somehow real. But they may be meaningful, useful, and provocative. As Marradi says, classification schemes and typologies “do not make assertions and therefore cannot be judged true or false”; rather, they may only be judged more or less useful for the purposes of this research.

3.3. QUALITATIVE CONTENT ANALYSIS

A second phase of empirical study picked up where the typological work left off: while the first phase was broad in scope, the second delved deeply into a small set of thematic research collections that may be considered representative of the breadth of the genre. A qualitative content analysis of three thematic research collections aimed to evoke – as thoroughly as possible – their characteristics, their commonalities, and their differences.

3.3.1. Purpose

A first phase of typological work gave me an aerial view of the landscape of the genre. This second step of empirical analysis zoomed in on select collections using a detailed qualitative

content analysis. Content analysis aimed to identify characteristics of representative collections, which distinguish thematic research collections as a genre.

Qualitative content analysis is a method for the systematic description of the meaning of a qualitative dataset. It aims to reduce data to those pieces or aspects that relate to or respond to an overarching research question (Schreier, 2013; Zhang & Wildemuth, 2009). In our case, that question is, what are the defining features of thematic research collections as a scholarly genre?

Given this question, the analysis produced thorough descriptions of three collections in terms of a set of attributes (or characteristics or properties). The set of attributes derived from a survey of the existing literature on collections, digital humanities evaluation and best practices, and alternative scholarly communication.

Qualitative content analysis is a common approach to a diversity of LIS research questions (White and Marsh, 2006). It is frequently the method used to interpret findings of interviews and focus groups, or to ask questions of relatively small textual corpora, such as journal runs or online discourse. In applying the method to thematic research collections, we treat the collections as documents, in keeping with our understanding that they are scholarly publications. Although content analysis is predominantly applied to texts, the raw material of this method is any communicative material, and its application to images is demonstrated (White & Marsh, 2006): “Another key factor is that the data communicate; they convey a message from a sender to a receiver. Krippendorff’s definition expands text to include ‘other meaningful matter’ (2004, p. 18).” This is imperative to our study, because thematic research questions are highly heterogeneous in terms of kinds of content, and are frequently multimedia publications.

3.3.2. Overview

The first step of this study was to collect a set of different potential attributes or characteristics of digital collections. This is a step toward answering the overarching research question: what are the defining features of the genre? The set aims to represent, as completely as possible, attributes derived from existing studies and literature on collections, publishing in the humanities, and evaluation and best practices for digital humanities projects.

The second step was to derive from those found attributes a set of categories through which we can analyze collections and how they operate in greater depth. This is a variation on the qualitative content analysis method (Zhang and Wildemuth, 2009), which is generally focused on systematic interpretation of texts. I adapt that process to a systematic study of the multiple media,

models, and tools that constitute collections. Conventional content analysis begins with the construction of a code book: a set of categories, with definitions and examples, which frame the analysis. Development of a coding frame relies in large part on inductive reasoning, but in this case was “directed” (Hsieh and Shannon, 2005) by existing sources of likely information about thematic research collections – the sources described below. The code book was iteratively refined as it was applied to the objects of study. The analytic protocol for this study, described below, is analogous to a code book.

The analytic protocol derives from a broad review of existing literature, to ensure analysis reflects the broad range of current thought and practice among experts and practitioners in relevant fields: the humanities, digital scholarship and publishing, and library and information sciences. Because the boundaries between collections and other kinds of digital resources and projects in the humanities are often indistinct, and because the genre continues to evolve and experiment at its own edges, I did not limit my sources to those specific to collections. In casting a wide net for sources, and in liberally identifying any potential aspects of collections that they mention, I sought to ensure that the categories used for analysis represent collections as completely as feasible.

The protocol was derived from 27 sources on the following topics: alternative or multimedia digital publishing (e.g., Ball, 2012); conceptual and empirical literature on digital collections (e.g., Palmer, 2004; Flanders, 2014); collection description schema (DC-CAP); and evaluation guidelines and recommended practices for digital humanities projects (e.g., Bates, et al., 2006; DHCommons;²⁹ IDE, 2014;³⁰ MLA Guidelines;³¹ Rockwell, 2012; etc.). For a complete list of sources, see the full protocol in Appendix B.

From these sources, I identified approximately 150 potential aspects of digital collections, including the types and extents of items they gather, how they may be navigated and searched, their functionality, and their underlying data models. There was significant conceptual overlap among attributes discovered across the literature, even if they were named or described differently depending on their authorship or the context of the source. Where overlap was discerned, categories were combined.

²⁹ <http://dhcommons.org/>

³⁰ <http://www.i-d-.de/publikationen/weitereschriften/criteria-version-1-1/>

³¹ <https://www.mla.org/About-Us/Governance/Committees/Committee-Listings/Professional-Issues/Committee-on-Information-Technology/Guidelines-for-Evaluating-Work-in-Digital-Humanities-and-Digital-Media>

For example, while one source asks, “Is there a legible intentionality behind the structure of the data?” (NINES/NEH³²), another asks the closely related questions, “Is there a clear statement of the standards that have been used, and an explanation of their benefits and/or limitations? Have the data been well constructed?” (Bates et al., 2006). These questions, and the collection attributes they imply, were combined with several others under the more generous category of “data models,” to be discussed below. This example illustrates another way in which aspects of collections gleaned from the literature were refined into analytical categories: by discarding what is prescriptive or normative in their description, and drilling down to characteristics at their cores. The aim of this analysis, after all, is not to evaluate the degree to which collections adhere to evaluation standards or best practices, but to make headway on the more fundamental questions of what these things are and how they work.

Finally, the final list of 38 categories was organized into 3 clusters, indicating thematic relationships between groups of categories: Context, Content, and Design. These groupings are intentionally loose. There is essential overlap within and between the clusters, and beyond that there are important relationships of other sorts obtaining between categories in different clusters, including dependencies.³³ The clusters are intended to give the set some organization for approachability rather than represent any ontological commitments.

Table 2 gives an overview of the analysis protocol. The full protocol is Appendix B.

Table 2. Overview of content analysis protocol

Cluster	Categories of analysis
Context	Theme; Purposes; Impact; Creators; Audience; Documentation; Provenance; Related collections; Related projects and publications; Review; Funding; Developmental stage; Host; Rights; Sustainability and preservation plans; Method
Content	Items; Diversity; Size; Narrativity; Quality; Language; Completeness; Density; Spatial coverage; Temporal coverage; Interrelatedness
Design	Data models; Navigation; Infrastructural components; Interface design; Interactivity; Interoperability; Openness; Identification and citation; Modes of access and acquisition; Accessibility; Flexibility

³²<http://institutes.nines.org/docs/2011-documents/recommendations-for-chairs/>

³³ These relationships may be worthy of exploration in future research.

Categories within the “Context” cluster pertain to a collection’s motivations, impacts, and other context, including where it came from, how it relates to the landscape of extant scholarly projects, and the provenance of its items (or data). The “Content” cluster includes categories about the nature and extent of what a collection contains. The “Design” cluster holds categories pertaining to the technical design of collections. Together, these categories represent the potential characteristics of digital research collections in the humanities, as suggested by relevant literature.

I then identified three collections for content analysis. Sampling for this method was purposive selection, aimed at informing the research questions under investigation (Zhang & Wildemuth, 2009). Three collections were chosen to represent the three central “types” identified in the first phase of typological analysis (discussed at length in 5.2):

Provisional-type 1 (collections provide direct access to primary sources along with advanced markup): *The Shelley-Godwin Archive*. This collection offers digitized, transcribed manuscripts from the influential Shelley-Godwin family of 18th- and 19th-century writers, including Percy Bysshe Shelley, Mary Wollstonecraft Shelley, William Godwin, and Mary Wollstonecraft. A substantial and still growing body of Shelly-Godwin manuscripts – including major works such as *Frankenstein* (M. W. Shelley) and *Prometheus Unbound* (P. B. Shelley) – appear both as digitized page images and as encoded texts. Manuscripts are supplemented with biographical, bibliographical, and other secondary sources. The purposes or intended contributions of this collection to scholarship include: providing unified access to related digital manuscripts that are scattered across a few collections; providing high-quality diplomatic transcriptions, with encodings that also highlight different authors’ hands on the same manuscript; providing flexible views and multimodal access to primary sources; and facilitating collaboration and curation.

Provisional-type 2 (collections also provide direct access to primary sources, but these collections afford minimal markup for various reasons): *The Vault at Pfaff’s*. This collection gathers primary and secondary sources about the historically significant bohemians of antebellum New York, U.S.A., particularly the large group of people “who were connected to the bohemian scene at Pfaff’s,” the historical restaurant and saloon that became an epicenter for a literary movement in the U.S. The site makes searchable an annotated bibliography of more than 8,000 texts by and about the “Pfaffians,” linking to

full-text primary sources both internal and external to the site wherever possible. The site also provides secondary sources, including a map, timelines, biographies, and historical accounts. Its purposes or intended contributions to scholarship include: facilitating unified search across a group of related people and the works variously associated with them, which are scattered across several digital collections; providing original, digitized page images of several influential periodicals of the era; identifying relationships among historically significant people and groups, and drawing connections from people and groups to texts.

Provisional-type 3 (collections provide indirect or mediated access to primary sources): *O Say Can You See: Early Washington, D.C., Law and Family*. This collection gathered, digitized, and analyzed freedom suits filed in Washington, D.C., and surrounding areas between 1800 and 1862, in order to explore multigenerational family networks, and the web of legal and social relationships that surround them, in early Washington, D.C.

The goal was to systematically analyze the contexts, contents, and design of these collections, to understand what they aim to do as exemplars of different *types* and how they go about it. By choosing representatives of types, I hoped to ensure that the collections I put under the microscope were sufficiently different from one another – in objectives, form, and content – that the whole analysis would not succumb to an overly limited or self-reflexive picture of what exists. Collections chosen were primarily in English (so that I could understand them); came from the same sources used for the typological sample (for the same reasons given in 3.2.2); and were openly accessible (so that I could freely assess them). The following remaining criteria guided my selection:

- Collections are well established: they are not in the earliest stages of development, they reveal some intricacy and purpose, and they do not show signs of deterioration (e.g., broken links, which would impede my analysis). They have been in active development for at least a few years. Two of the collections are fairly young but well established. The other, Vault at Pfaff's, is an older project but continues in active development.
- Collections are well documented. There is a great range in the extent and quality of documentation for thematic research collections. It turns out that, for the most part, provisional-type 1 collections are documented better than most, both in terms of

technical documentation and editorial decisions, and even the provenance of the data itself. This is probably because most of those adhere to the TEI guidelines and grew up within the text-encoding community, even prior to TEI, both of which afford space for and encourage documentation. There is some variation in the strength of documentation even among the collections I chose to study, but they are relatively more transparent in their technical and editorial choices than most thematic research collections. I did not choose to correspond with collection creators in order to augment what is publicly available about them. This would have added a burdensome human-subjects element and seemed unnecessary in light of the extent of publicly available documentation.

- Collections are complex. The reason for this priority is, first, that they are simply more interesting to spend a lot of time with. It is also important to study complex collections because they pose the greatest challenges to our existing systems of collection, preservation, discovery, access, and sustainability. As such, they do not fit readily into the mold of a simple content management system or institutional repository.

3.3.3. Limitations

As an inherently reductive approach, which relies heavily on the notion of *category*, content analysis is well suited to the identification of distinct features. This kind of feature-based approach to understanding a genre or a resource is firmly rooted in the epistemological traditions of our field, particularly in classification and cataloging. Indeed, there is resonance between classification as a method and qualitative content analysis; the final products of qualitative content analysis are usually descriptions or typologies (Zhang & Wildemuth, 2009). To broach “defining features” we required a more thorough, descriptive and interpretive study of aspects of collections that were, by necessity, treated more reductively in the provisional typological work. However, I acknowledge that pulling at threads is bound to yield a limited view of the whole.

I have mentioned that the coding frame will begin with properties already known or hypothesized to be endemic to thematic research collections. However, the genre is still developing and often experimental; there are bound to be exceptional cases that do not conform to any results of this analysis. This study is not designed to generalize across all thematic research collections but rather to establish a set of defining characteristics that expand upon our existing characterizations of thematic research collections, and inform our practical decisions about their development and treatment.

Qualitative content analysis often asserts its rigor by specifying a level of agreement obtained between multiple analyzers or “coders” of a dataset. Inter-coder reliability is impossible to measure in a solo study. This is an acknowledged limitation of this work. To appease the dangers of proceeding alone, the coding frame will aim for clarity of description of categories and will thoroughly describe criteria for decisions made in the application and description of codes.

3.4. INTERVIEWS

The third phase of this study turned from direct contemplation of thematic research collections to an interrogation of aspects of their systemic context. I conducted a set of interviews with representatives of library publishing programs and digital humanities centers, with the aims of describing current practice around thematic research collections in libraries and related scholarly-publishing entities, and revealing challenges to their integration into library systems of collection, discovery, access, and ongoing maintenance.

3.4.1. Purpose

A set of semi-structured interviews with nine practitioners revealed challenges to supporting the genre, and particularly challenges to sustaining and preserving thematic collections over time.

Interviews will address the following overarching questions. The first question pertains to the generation of collections. The second pertains to their ongoing usefulness – specifically, it evokes primary duties of the library toward its collection.

- How do library publishing programs and other scholarly-publishing entities support the creation and publication of thematic research collections, and what problems exist in meeting the needs of scholars and collection creators?
- How do libraries collect, represent, describe, preserve, and otherwise treat thematic research collections after publication, and what problems exist in meeting the needs of potential user communities?

The goal of this phase of the study was to produce a descriptive account of how thematic research collections are created and handled, and a sense of the challenges to and opportunities for library collection (including description, representation, access-provision, and perhaps preservation), which may lay a foundation of understanding for ongoing research and perhaps eventual, normative or best-practice recommendations. The interviews provided supporting

evidence for outcomes of the typology and content analysis, as discussed in Chapters 4 and 5. Chapter 6 details the central outcomes of the interviews: challenges, strategies, and roles in the sustainability and preservation of collections.

3.4.2. Overview

Sampling for this phase of the study was purposive. I selected participants most likely to know the most about this genre and its systemic contexts. I prioritized the potential richness of expert response over any gains in generalizability that might be attained from some kind of random sample. I wanted the results to be representative, not of a population, but of the state of the art of the publication of thematic research collections.

Sources for the earlier phases of this study (typology and content analysis) served again as sources for finding potential interviewees. Participants were selected to represent the main institutions that provided the sample of thematic research collections analyzed through typology and content analysis, namely the Center for Digital Research in the Humanities at the University of Nebraska-Lincoln, the Maryland Institute for Technology in the Humanities, the Roy Rosenzweig Center for History and New Media at George Mason University, the Scholars' Lab at the University of Virginia Library, and the Institute for Advanced Technology in the Humanities (University of Virginia),, determined by the sample identified for the typology and content analysis. Where possible, I interviewed more than one person from each institution. Two additional interviewees were selected for their extensive experience working with collections, and their expertise in library administration. The participants all waived confidentiality. Table 3 lists the participants in alphabetical order by last name, and for each gives a participant ID used throughout the rest of this dissertation for readability, except in cases where a name is necessary to contextualize a quotation or anecdote. The table also gives their affiliations and relevant positions at the time research was conducted.

Table 3. Interview participants

Participant ID	Name	Affiliation	Position
P1	Jeremy Boggs	Scholars' Lab	Head of Research and Scholarship
P2	Neil Fraistat	MITH	Director
P3	Andrew Jewell	CDRH	Professor of Digital Projects
P4	Sharon Leon	RRCHNM	Director of Public Projects
P5	Worthy Martin	IATH	Co-Director and Associate Professor of Computer Science
P6	Trevor Muñoz	MITH, University of Maryland Libraries	Associate Director and Assistant Dean for Digital Humanities Research
P7	Bethany Nowviskie	Digital Library Federation at the Council on Library and Information Resources	Director
P8	Brian L. Pytlik Zillig	CDRH	Professor and Digital Initiatives Librarian
P9	John Unsworth	University of Virginia	Dean of Libraries, University Librarian

Thematic research collections are spawned in all kinds of contexts, but most often in digital humanities centers. Therefore, the interviews will begin with people at these kinds of institutions who have experience in helping scholars develop these publications. If the same people were in the position to attest to the ongoing management of these resources, especially in the library context where applicable, the interviews continued with them onto questions of management and maintenance. Otherwise I sought their assistance in locating secondary respondents at the same institution.

As systems of digital scholarship are structured differently at every institution, and because libraries play different kinds of roles in those systems, the protocols guiding the interviews were necessarily very flexible, and tailored to participants' affiliations, positions, and expertise. Appendix C gives the basic interview protocol, which was adapted to guide each interview. Interviews lasted approximately one hour. Most were conducted over the Skype or phone. One

was done in person. Once interviews were complete, I transcribed them, and then subjected them to qualitative content analysis (Zhang & Wildemuth, 2009).

Interviews were coded using qualitative content analysis (see full-blown discussion of qualitative content analysis in section 3.3, above). The coding frame was built inductively, deriving categories (themes) from the transcripts in answer to the research questions (Zhang & Wildemuth, 2009). Categories covered the following topics:

- audiences/users
- collaborativeWorkflows
- collectionChange
- collectionRelationships
- concept/genesis
- culturalHeritage/publicHumanities
- data
- design/dev
- discovery
- documentation
- experimentation
- flexibility/mobility
- genre
- impact/evaluation
- libraryCollection
- libraryDescription
- processVsProduct
- purposes
- review
- roles
- scholarly communication/publishing
- sustainability/preservation

Some of these categories arose from the research questions; others emerged as unexpected themes from the interviews. Once the transcripts were fully coded, I analyzed themes for meaningful and potentially significant answers to the research questions. Results – relevant to collection purposes, sustainability and preservation, and flexibility of collections – primarily appear in Chapters 4, 6, and 7.

3.4.3. Limitations

Interviews in general are limited by subjectivity. The results will not be generalizable to the whole population of people in DH centers, libraries, or other institutions that are actively or potentially engaged with thematic research collections.

Few programs are explicitly or visibly working with these kinds of collections in any systematic way. Due to rarity, this study aims to be more exploratory and foundational than comprehensive or conclusive about the research questions. In particular, few libraries appear to systematically deal with thematic research collections post-publication. I found that there is sparse extant knowledge on how libraries do or can deal with thematic research collections after they are created. That in itself would constitute an important finding, but in that case, the study has pursued an understanding of their current management and maintenance outside of the library.

CHAPTER 4: COLLECTION PURPOSES

4.1. INTRODUCTION

This chapter considers Research Question 1: What are the defining features of thematic research collections as a scholarly genre? What features are common to thematic research collections? What features distinguish thematic research collections from other kinds of collections?

It seems impossible to talk about the defining features of collections without talking about what motivates them: the many distinctive and changeable purposes of thematic research collections. Through typological analysis, content analysis, and interviews with practitioners, it became clear that collections are distinguished in part by what kinds of contributions to scholarship they intend to make. Intended contributions, or purpose, set collections apart from other kinds of scholarship, and indeed from other kinds of collections.

We have long understood that digital scholarship aims at different targets than do conventional scholarly products. It is a theme in the digital humanities literature that digital scholarship should aim beyond the capabilities of print (though toward *what* is usually left vague). In peer-review guidelines for digital humanities work, Bates et al. (2006) ask about the purposes of digital resources:

Does the material contained in the resource benefit from having been made available digitally rather than (or in addition to) in print? Have the resource creators considered a sufficiently wide range of uses beyond print? Is it important that digital presentation should add value, or is it simply enough that the material is made available at all?

Similarly, Thomas (2016) laments the failure of 20 years of scholarship to produce much indispensably digital, interpretive scholarship: “there were few hypertextual works that embodied complexity³⁴ or altered the mode of scholarly communication in ways *uniquely suited* to the online space” (emphasis added).

While the literature seems united on the notion of the transcendence of scholarly purposes in the digital realm, there is no common sense of what the purposes of different digital genres may

³⁴ Here Thomas is drawing on Ayers (1999), who described the future of hypertext historical narrative as, “embody complexity as well as describe it, to permit the reader some say in how history is conveyed, to create new spaces for exploration”. Ayers seems prescient: the notion of *embodying complexity* in order to allow opportunities for exploration and interaction is central to the purposes of thematic research collections.

be, or what they aim to contribute to humanities scholarship. Most of the peer-review guidelines and best practices literature reviewed in the course of content analysis suggests that purpose is unique to each digital creation. For example, Anderson and McPherson (2011) suggest that each scholar is individually responsible for explaining “the unique contributions of a work...[and] the most useful ways of assessing influence and quality.” It is likely that we see little consistency in the literature about the exact purposes and functions of digital scholarship because they change rapidly, and because those involved with creating digital scholarship are immersed in experimentation and boundary-pushing. Nonetheless, calls for genrefication grow louder, in part because identifying genres with recognizable purposes makes the processes of scholarly evaluation and communication more efficient.

Studying the purposes of thematic research collections in depth has the potential to help those who develop and evaluate scholarly products establish a shared sense of intentions, building on Flanders (2014): “By identifying a genre or a set of scholarly practices through this nomenclature, we are also saying something about our own intentions ... a specific interpretive end or set of research goals, a specific kind of epistemic outcome.” The goal of this chapter is to examine and elaborate the various purposes of extant thematic research collections.

What are the purposes of thematic research collections, and how do they differ from other genres of scholarship? This section expands on the purposes identified in Palmer (2004), focusing on the varieties of purposes raised by interview participants and emerging from the content analysis of collections.

Not only do thematic research collections diverge from conventional products of humanities scholarship, they reflect less of traditions of collection in libraries and archives than might be supposed. As Flanders (2014) cautions, thematic research collections often resemble (and may be based in) physical collections, but that belies “the novelty of digital collections and the distinctive epistemological conditions under which they present themselves to us.” Library collections are governed by institutional mission, rather than specific research objectives; archival collections base their collocations on originating source (Palmer, 2004). By Palmer’s account, thematic research collections, on the other hand, aim to:

- Bring together a thematically coherent yet diverse set of sources to support research (*collocation and access*)
- Support specific activities with tools and functions for discovery, reading, annotating, comparing, linking, mapping, modeling, etc. (*activity support*)

- Manifest collection-creators' research and interpretative advances (*interpretation and analysis*)
- Facilitate collaborations between researchers across time, space, and disciplinary lines. (*generativity*)

This study confirms and expands upon the first three in the corresponding subsections of “Foundational purposes” below: collocation; infrastructure and activity support; and interpretation and analysis. The fourth purpose, generativity, is given a thorough treatment in section 4.3, because it adds new dimensions to our understanding of what and how collections contribute. Because purposes are inseparable from the perceived audiences served by them, this chapter finally turns to an exploration of the diverse audiences for thematic research collections.

4.2. FOUNDATIONAL PURPOSES

Typological and content analysis and interviews all suggested that collections serve a multiplicity of different, sometimes competing purposes. Section 3.3 briefly described the goals of the three collections studied in the qualitative content analysis. I list collection purposes in greater detail here just to illustrate the array of goals toward which collections are directed. Content analysis revealed a variety of explicit purposes motivating these projects; there may be more unstated goals (see Table 4). In contrast, the central purpose of traditional publication in history may be described as, “demonstrating an extensive, closely reasoned argument within a larger interpretive framework,” and linking it to evidence (Harley et al., 2010).³⁵

Collections may hold multiple purposes at once, but collections may also become reoriented to different purposes over the course of their lives. Some purposes change over time; others remain static. This is true for individual collections, and it may also be true for the genre as a whole.³⁶ Participants described how their ambitions for their collections shifted in relation to numerous factors: realization of original goals, changing funding sources, levels of support, staffing changes, changes in copyright status of original sources, technological enabling factors, outcomes of experimental and development efforts, collaborative workflows, changing

³⁵ Harley et al. (2010) do identify numerous other purposes for publication, but those operate at a different level of granularity than the purposes identified in Table 4, in part because they were gleaned through interviews and focus groups rather than content analysis. Those purposes included things like staking claims to research ideas, bolstering one's own reputation, evaluating others' work, sharing evidence, etc.

³⁶ A few participants suggested a vague sense of there being different historical “eras” or “epochs” of digital-collection-development or digital-humanities scholarship, generally. This may be a direction for further inquiry, if we want to understand more about the history of the genre's development.

institutional contexts, shifting standards and best-practices, changes in perceived audiences, and simply the generation of new ideas. One participant reflected on one long-running project:

there had been many different groups and models for how the work happened at the university, and there's just a lot of, again, that coral reef feeling – it had all grown up organically and it was all done in different time periods, to different standards, with different understandings even of what the goal was –P7

Table 4. Collection purposes

Shelley Godwin Archive	Vault at Pfaff's	O Say Can You See
<ul style="list-style-type: none"> • Collocate a complete set of manuscripts • Digitize manuscripts as high-quality page images • Transcribe and encode manuscripts • Facilitate multimodal reading and exploration • Expose scholarly texts to wider audiences • Experiment with crowd-sourced and distributed encoding • Facilitate participation, annotation, curation 	<ul style="list-style-type: none"> • Aggregate access to distributed, related content • Illuminate a network of people, places, and texts • Digitize unique periodicals • Facilitate discovery, inferencing, and impact assessment about historical figures and works • Provide biographical and bibliographic context for items and people • Serve teaching and research purposes • Publish unique secondary sources 	<ul style="list-style-type: none"> • Digitize archival documents for broader access • Aggregate legal and other records as evidence • Transcribe and encode documents for analysis • Analyze family and social networks • Expose hidden relationships and histories • Generate a social network, including family trees • Provide numerous access points to the historical record • Publish historical essays illuminating the collection

There is also a distinction between, and sometimes tension between, a collection's short- and long-term goals. Most collections are funded by grants, fellowships, or other short bursts of support. Long-running collections are more or less successful at leveraging intervals of support toward overarching ends. For example, some participants described adopting certain data models in hopes of one day supporting imagined functionalities or uses of the collection. Others were less prescient, and described having to pursue funding to remodel and rebuild collections in light of technological shifts and shifts in purpose. One participant described trying to balance between short- and long-term priorities, in the course of development and data modeling:

I think it's sort of like what Stephen Sondheim used to say about writing songs³⁷: which came first, the words or the music? I think he said something like, you shouldn't let the music get too out in front of the lyrics and vice versa. You need to keep developing them together. ... I still think it's really important not to focus too much on your data model: you don't want to get stuck in this vacuum of thought that you regret a year or two later when you're trying to leverage this data, or handle this data, or manipulate it in some way. –P8

Purposes may be implicit or explicit. One content-analysis source asked, “What does the [resource] promise explicitly? What does it merely suggest by self-classification (e.g. ‘edition’, ‘critical edition’, ‘portal’, ‘collected works’, ‘digital archive’, ‘virtual archive’ etc.)?” (Sahle, Vogeler, and IDE, 2014) And sometimes the specific purposes of collection development are vague or invisible to users, as Palmer (2004) notes: “While a loose theory of collecting may be guiding creators’ selection of content, the criteria being used to determine what is appropriate for a collection and the long-term development principles of a project are not always clarified for users.” Indeed, collection creators may not always be fully aware of their own purposes, particularly when they are motivated by experimentation. Purpose can be especially tricky when it is oriented toward unknown uses, as most collections are. P1 expressed a common theme in the interviews – that collections inevitably aimed to support hazy or unknown kinds of use: “I don’t have the answers for ...how people have used those sorts of collections and what they actually use them for.” We will come back to this theme under “Audiences”, below.

The first few of these purposes are familiar; aspects of them are even entailed by the very concept of “collection.” I rehearse them here, in part because the interviews and content analysis have enriched my understanding of these aspects, and in part to lay a conceptual foundation for what follows.

4.2.1. Collocation and access

The most common and fundamental purpose of thematic research collections is to gather sources for research together. There are a number of ways to accomplish this. Some collections, in fact, do not gather primary sources themselves but unify access to them through aggregated metadata. Collections gather items to support deep inquiry on a particular theme (Palmer, 2004).

³⁷ Indeed, Stephen Sondheim said almost exactly this: “...you can paint yourself into a corner if you write a whole tune or even half a tune with no idea what you're going to say in it...”, quoted in <http://www.npr.org/2012/02/16/146938826/stephen-sondheim-examining-his-lyrics-and-life>

Virtual collocation serves to subvert original organizational contexts of items – usually library and archival organization – which are each structured around different organizing principles. The principle that guides collection development in a thematic research collection is contextual mass (Palmer, 2004), which may be understood as a confluence of thematic cohesion, completeness, and rich context. What a thematic research collection gathers is (at least) a set of primary sources, which constitute potential evidence for different potential lines of inquiry. The implication of “evidence” is that the sources must be authentic (to the extent that they can be) and of sufficient quality that they are fit for whatever interpretive or analytic purposes they are intended. Lest we take for granted the significance of collocation, one participant testified to the power of simply bringing sources together in one place:

the first time I used the ‘compare’ feature on the Blake Archive, I – and I’m not kidding – I wept. Because there I could see across my screen like eight different instances of the same plate in all their variation. I thought about all the generations of Blake scholars before me, who had to rely on memory and notes and the insufficient ways that you can keep things in your head when you’re trying to think across a field of difference. So from the first I saw the powerful ways – beginning just in terms of access and search – that these archives could start enriching scholarly lives –P2

The act of collection is a first step in a lot of humanities research processes (Brockman et al., 2001), and it serves the purposes of the individual researcher. The subsequent step of publishing a collection on the web serves other purposes – and the first is to provide other people with access to the collection of sources.

4.2.2. Interpretation and activity support

Beyond collocation and access of primary sources, the next most common general purpose of thematic research collections is to support scholarly activities, including searching, reading, annotating, comparing, referring, selecting, linking, and discovering (Palmer, 2004). This is often accomplished by installing affordances for various kinds of use of the primary sources. Those affordances, in turn, both express and facilitate analysis and interpretation to different degrees. The affordances of thematic research collections are numerous. Table 5 gives a list of secondary sources and tools encountered in the initial typological survey of 98 thematic research collections.

Table 5. Secondary sources and affordances identified in survey of 98 thematic research collections

3D models and digital reconstructions	exhibitions and exhibits
annotation tools	explicative, historical essays and stories
bibliographies and annotated bibliographies	family trees and kinship diagrams
biographies	glossaries
blogs	indices
catalogs	interviews
chronologies and timelines	journals
codicological descriptions	maps and cartographic tools
collation tools	network analysis
collection lists	prosopography tools
commentary, reviews, and critical resources	statistics
concordances	teaching guides
creative works	text analysis tools
discovery tools	transcription tools
documentaries	translation tools
documentation	visualization tools and visualizations

Affordances serve to define how collections can be used, and they mediate access to items in different ways and to varying degrees. Some intervene between users and items directly, as in the case of analytical tools or visualizations through which items are manipulated, or narratives into which items are embedded. Others interpose less directly, as in the case of discovery tools and navigational facilities, which nevertheless influence how items are encountered and the contexts in which they are understood. Most collections make primary sources directly accessible as such via search and browse, but some do not. This is discussed in greater depth in the next chapter.

Interpretation may also be embedded in the data models underlying collections, as many have recognized. Martin described embedded interpretation and affordances as the “connective tissue” of the collection, arguing that it can be richer and more expressive than traditional publication. Of course, embedded interpretation can also be invisible. Thomas (2016) described this phenomenon as endemic to the last 20 years of digital humanities scholarship:

Scholars built digital archives, layered them with affordances that were premised on interpretive decisions, then wove interpretive scholarship into a digital project. So interwoven were these activities that non-digital scholars could see little that resembled their expectations for peer-reviewed scholarship. –Thomas, 2016

Embedded interpretation and activity support constitute central purposes of thematic research collections, and they also reflect the genre’s *hybridity*, which is usually discussed in two senses:

- First, the genre is both publication and platform, both product and ongoing process, an expression of research and a tool for further research. As Palmer (2004) says, “research takes place in the production of the resource, and research is advanced as a result of it. Thus, scholarship is embedded in the product and its use.”
- Second, collection purposes may “blend, redefine, or render obsolete the traditional boundaries between teaching, research, and service” (Modern Language Association, 2012). This is something I come back to under “Audiences,” below.

It is this hybridity of purpose that can serve to either valorize or compromise thematic research collections as scholarly products. Leon observed, “If we approach [collection-creation] in a way that doesn’t recognize it as an act of scholarship, we’re being willfully ignorant about the conditions of creation that shape our evidence base.”

4.3. GENERATIVITY

In a basic sense, collections serve to generate meaning. The purposes described above – collocation, access, interpretation, activity support – serve both creators and users by helping to determine, reveal, or construct significance, purpose, underlying truth, import, and implication of items in relation to one another. One participant reaffirmed the meaning-making potential of collections:

...we had always understood the Rossetti Archive...to be this *complex, fungible machine for meaning-making* about Rossetti and his corpus. And really understood all thematic research collections or digital archives to be that at their heart” –P7, emphasis added

Collections are generative in a number of ways. We will explore the generative potential of thematic collections at length in the final chapter of this study, on “Collections as platforms.” Here I enumerate the kinds of generativity described by interview participants, and which expand upon Palmer’s (2004) account of collections as platforms for inquiry and collaboration between researchers across time, space, and disciplinary boundaries.

4.3.1. Experimentation

Collections become loci of experimentation and innovation, so that they generate insights about things even beyond the themes they are built around. They aim to help advance research on new techniques, tools, and analytical methods, which may take the shape of affordances built on top of the collections' items, but which – more importantly – are intended to have broader application to other collections of sources, or to other kinds of digital scholarship across disciplines. Several participants described collections almost as laboratories for development and innovation of tools, techniques, and methodologies.

For example, in an effort to make the Rossetti Archive more flexible for use by scholars from different disciplines, Nowviskie described how the project team developed a prototype exhibit-builder, which would allow, for example, art historians and literary scholars each to create different, custom portals into the collection. This exhibit-builder became Collex, which in turn spawned Project Blacklight, which became a popular discovery layer for digital collections generally. Nowviskie reflected:

We always loved that aspect of it – it was digital humanities scholars trying to solve a problem for themselves, related to their own scholar-built collections, that ultimately wound up solving some problems for the library writ large.

Ongoing innovation is imperative to collections' sustainability; as P2 observed, “one thing you can't do is stay still, obviously.” One participant described how continued experimentation and innovation become central to the purposes of thematic collections as projects, once the first purpose – of gathering content – is achieved:

Once you've got the strong content underneath it, then the real quest is to figure out how to make it available and increasingly useful in compelling ways. We have this ...terrific electronic edition of *Frankenstein*, and there is a group of DHers in the Pittsburgh area...[who] are in the midst of a number of experiments with the data to see what they can learn. And that will get refracted by way, not just of the outcomes, but illuminations of their methods – you know, how did they do that? Why did they do that this way? –P2

Other interview participants confirmed the centrality of experimentation to the lives of their collections. Collections do not only generate questions about their subjects; they serve as sandboxes for generating and testing methodological, technical, and representational questions. One participant expressed his (qualified) enthusiasm for these kinds of outcomes of collection-

making: “All these things are kind of black holes of fun you could fall into. ...The problem is it’s impossible to know at the outset exactly how difficult it’s going to be” (P8).

4.3.2. Collaboration

Collections exist in part to serve as focal points for collaborations, both among their creators and among their users. This is related to the notion of experimentation; most experimentation depends on sometimes extensive, distributed collaborations. And it happens both on the back ends and front ends of collections.

Speaking of the Walt Whitman Archive, Jewell observed, “a thematic research collection isn’t just a product. It’s also a group of people working together who want to try new things. It becomes kind of like a unit, a little academic unit within the big institution.”

Jewell described a conscious effort to foster a culture of experimentation, incorporating “a broad array of collaborators,” through the Cather Archive:

a lot of the novel affordances, making those happen, has to do with the culture of the project internally and whether you allow for some balance of efficiency and productivity and experimentation, including things that could fail...

The notion of constructing an environment that accommodates the risk of failure came up in more than one interview. Fraistat described the omnipresent potential for failure in the development of Romantic Circles: “We were going to be trying out stuff and we didn’t know if it would work, what would work, what wouldn’t” (P2). He described how experimental affordances might struggle or fail at one time, but be recuperated later as a collection’s user community grows and technological developments enable new modes of implementation. For example, Fraistat described trying to implement a forum for substantial and immediate scholarly exchange in Romantic Circles, which failed in its original incarnation because scholars were unwilling to publish substantive arguments in an informal medium. However, a similar functionality of the site has been recreated as a “Reviews and Receptions” section, which incorporates nontextual media and live, synchronous hangouts – with which scholars are at this point more willing to engage. Something might fail not because of its actual or potential scholarly or intellectual worth as an affordance of the collection, but because of its timing, its particular implementation, or the current size and interests of the audiences for the collection:

...some of the ideas we had were good ideas. The forms in which they were manifested weren't always successful, or they were successful for a time, but we seem to find different ways to start recuperating them in new forms over time – P2

Collaborations, in turn, bear fruit beyond the outcomes of the collection itself. As hubs for innovation, interpretation, and other scholarly work, collections create research networks and opportunities for publication and reputation-building. Jewell described the Whitman Archive as “really generative” in the sense that many of his colleagues had been graduate students working on that collection, and noted that “it has been good for a lot of people in their careers.”

By serving as a locus of collaboration between faculty and librarians, collections have the potential to “change the way faculty think about what librarians do and have to offer” (P9). Several participants described collaborations around collections as aiming to help reorient the library toward research partnership over service-provision. Boggs, speaking from Scholars’ Lab, a digital humanities unit that is part of the library at the University of Virginia, reaffirmed the importance of partnership and collaboration: “we do try to be collaborators as much as possible, and not simply providing a service...People actually enjoy having collaborators [who] are genuinely interested in their research questions, and not simply providing a solution”. Another participant suggested that the whole mode of scholarship depends on dialog and interleaved intellectual contributions from faculty and technologists. Those collection-centered collaborations are necessary, not only to forward the immediate goals of collection-creators, but to help crystallize their overarching research questions and ultimate purposes:

there's that interaction, of trying to collaborate to try to understand what it is they want and flesh it out and make it clearer to them ... most of these thematic repositories of collections seem to require some notion of collaboration with the technologists, so that there's almost always something to bring to bear intellectually from the technology point of view –P5

4.3.3. New research directions

As a collection grows to suit each new purpose, accreting not only content but collaborative teams and infrastructure for innovation, it opens up new research possibilities, clarifies extant lines of inquiry, and reveals new and urgent questions. One important way in which collections are generative is that, in addition to generating insight and meaning, they serve spin off new projects,

new directions for research. Jewell described how different phases of the Cather Archive had combined to position them to create an authoritative scholarly edition of Cather's letters:

...in some ways this edition we're working on now is what I always wanted the Cather Archive to be, but it hasn't quite been before ... the value of the edition itself, of the scholarly edition as a scholarly product, that emerged in my consciousness through working on thematic research collections, and so therefore made the project I'm currently doing visible to me in a way that it wouldn't have been. And so that's a really good point, that work on a thematic research collection has generated this project that's within that

Two interview participants noted that collection-building served to help humanities scholars clarify and re-conceptualize their research questions and objectives, or reveal and open whole new lines of inquiry. Part of this is serving as a sort of hub for collaboration. Martin, who works intensively with IATH faculty fellows to design and implement collections, noted:

...a good number of the fellows, when they anecdotally respond to what was their main benefit of being a fellow, they'll talk about their own individual reconceptualization of what it was they were doing, that they came to a much clearer and possibly different perspective on the question they were actually asking, how to ask those questions.

One participant suggested that this kind of generativity, in fact, is an ideal of thematic research collections:

...in general I think there are two different questions these collections help answer. One is, what are questions that your peers in your field want to do research on that they can't yet without this collection being made available and public? And two, *what questions should these collections help provoke, that we haven't even thought of yet?* I don't think people should have the answer to that, but I think they should work toward being aware of that sort of thing. —P1 (emphasis added)

Boggs highlighted an example from Take Back the Archive, a University of Virginia collection intended to "preserve, visualize, and contextualize the history of rape and sexual violence at the University of Virginia, honoring individual stories and documenting systemic issues and trends."³⁸ While combing through the collection one day, a question occurred to one of the collaborators: Was there any aggregate record of University of Virginia scholarship on rape

³⁸ <http://takeback.scholarslab.org/>

and sexual violence? It turned out that there was no record or aggregation – within this collection or anywhere else – of extant, local research on these themes. This became a new direction of inquiry, and a new development priority, for collection creators. Boggs noted that if collection creators can become self-aware about gaps in the collection, those can be valuable indicators of opportunities for new research, new gathering, and new scholarship:

Lisa [Goff, faculty lead] had not thought about that question until encountering this archive....That's a good thing. I think people who make these collections worry that they haven't put in enough, or that it's a shortcoming if there aren't things in a collection, but I think if instead of being paralyzed by that or hindered by that, that that's actually a good thing and it's good to be aware of –P1

4.3.4. New evidence

Collections gather sources for research, but they may also serve to generate new sources and new forms of evidence. Some collections exist to collect *from* their users, gathering oral histories, personal stories, and digitized artifacts. This is discussed in greater depth below under “Audiences as co-creators.” Some collections that are not soliciting new, original sources may have a purpose of extracting or deriving data and other forms of evidence from their existing sources, to represent new primary sources. For example, the Bracero History Archive undertook a massive, distributed collecting effort in order to gather evidence, ranging from oral histories to digitized artifacts, from the public about the lived experiences of workers in the Bracero program.³⁹

Some thematic research collections are created to fill representational gaps in extant institutional collections, or gaps in prevalent histories or literary accounts. Both Vault at Pfaff's and O Say Can You See align themselves with this kind of purpose. Vault at Pfaff's does not predicate its selection or inclusion of items on their historical or literary value; it leaves that assessment up to users of the collection. Its goal is to “provide access to the primary and secondary source documents that will allow students of American culture to determine who the Pfaff's bohemians really were and to assess their contributions to the art and literature of the antebellum United States.”⁴⁰ Pfaff's exists in part to facilitate the discovery and exposition of undiscovered, potentially significant historical and literary connections about “an undervalued moment in American cultural history.” O Say Can You See explicitly aims to fill gaps in the historical record, both evidentiary and knowledge gaps. It aims:

³⁹ <http://braceroarchive.org/about>

⁴⁰ <https://pfaffs.web.lehigh.edu/node/38090>

...to make visible what has been invisible in the history of slavery, including the networks of relationships of the enslaved and free. Scholars have written accounts of slavery based on models that have been quantitative (economic), institutional, political, and cultural. ... It was also experienced in individual actions and individual movements through space and time, the traces of these largely invisible in the historical record – O Say

The purpose of a collection may be to provoke social or political change. The Take Back the Archive collection arose from a shared desire to raise institutional awareness about sexual assault at the University of Virginia:

“[Lisa Goff] started this archive to address what she saw or felt as an institutional or community amnesia about sexual violence at [the University of Virginia]. She was shocked at how many people ... were saying things like, ‘they didn’t know that things like this happen here,’ when in fact she had collections of things going back decades” (P1)

Public-history collections, thematic collections of primary sources that are more oriented toward use by the general public, can also serve to gather original sources. We return to this theme under “Audiences as co-creators,” below.

4.4. AUDIENCES

Thematic research collections appeal to broader audiences than other kinds of scholarly product, in part because most collections are openly accessible on the web, and in part because they can be put to varied types of use to meet diverse interests. Jewell described the choice between collection and the publication of a more conventional edition as a vehicle for the curated content of the Cather Archive:

There would’ve been a way to approach this same project very differently, and I think inferiorly: like a big, extensive, multivolume press edition, which would’ve had a very small, academic-library-centered audience. And that’s all. ...It would be more satisfying to me if somebody teaching high school juniors were able to use it in the classroom alongside professional scholarship

Of course, it is the physicality (and cost) of print publication that would limit the audience of a multi-volume edition. Beyond the fact of being digital and being openly available on the web, how do thematic research collections aim to serve broad audiences? The foremost way in which collections engage broad audiences is by acting as platforms for research. We will come back to

how collections serve as platforms for research and learning in the next chapter. But collections also engage audiences through affordances.

Collections add layers of interpretation and affordance that can pivot the collection of sources toward different potential uses and audiences. Once items of potentially broad interest are gathered, collections add affordances and modes of interaction that can reorient how different kinds of audiences encounter and experience the contents of the collection. For teachers and students in primary and secondary school, collections may offer teaching guides. Chronologies, interactive maps, and interpretive and analytical stories serve to enliven collections for teaching and learning at higher levels. For research users, collections provide advanced searching mechanisms, comparative viewers, and annotation functions. For example, the Vault at Pfaff's is self-conscious and explicit about how different components of its site highlight and expose the collection in different ways:

The *People* and *Bohemian New York* sections of the site are similar to *gallery exhibits* one might associate with public humanities projects in that their goals are not to provide an exhaustive repository of texts, but rather to introduce scholars, students, and interested readers to somewhat broader topics. ... We have conceived of these sections of the site as having more of a storytelling function, as opposed to the research function of the Works section of the site. —Vault at Pfaff's

These are all layers built upon the store of collected content, and each layer exploits the data models that have been put in place to afford multiplicity of potential uses.

The most common articulation of different audiences is the familiar division between teaching and research, but this is not the only division – and teachers, students, and scholarly researchers are not the only audiences. To varying degrees, collections also cater to different public audiences, including international audiences. Another potential audience that should not be overlooked is that comprised of the collection-creators themselves. Finally, there are unanticipated audiences, which nevertheless may influence collection design and development.

4.4.1. Public-facing collections

Many collections have potential research functions but are presented more immediately as “public humanities” projects, most often public history projects. For these collections, the primary audiences are different communities within the broader public. Within those communities, collections may appeal to different kinds of use, including general interest learning around heritage

and history, and non-scholarly but still intensive research uses by citizen humanists. We will come back to these below, under “Audiences as co-creators.”

Collections designed for scholarly audiences may reach out to public audiences to broaden their user communities and increase their impact. Jewell described upcoming events at the Cather Archive, in celebration of the 100th anniversary of Cather’s novel *My Ántonia*, which seek to engage communities throughout the state of Nebraska, where Cather grew up on the frontier:

...we’re trying to use themes from the novel to reach out into the communities, immigrant communities, environmental communities, Women at Work, ...to try to think about public humanities and the way it can be relevant to people who maybe don’t think of hundred-year-old books as relevant to them

For some collections, the effort to appeal to broader audiences may be necessary to retain the kind of community engagement that leads to sustainability. For example, Muñoz notes of the Shelley-Godwin manuscripts, “The community of scholars that needs that detailed a look at the manuscripts, how a text evolved, is necessarily small because it’s a very specialized inquiry. But [the collection] is uniquely valuable to that population.” It turns out that the Shelley-Godwin archive found a significant, public audience, mostly due to its popular centerpiece, *Frankenstein*. In fact, the Shelley-Godwin team purposely and, in Fraistat’s own words, “shamelessly” launched the collection on Halloween, to take advantage of mainstream media’s annually recurring interest in *Frankenstein*-related stories. The strategy was effective. The site saw 60,000 unique visitors in the first 24 hours. The project aims to continue to engage public audiences with events like “Frankenreads,” an “international day of public reading of *Frankenstein* at local libraries, at major research libraries, at museums, at high schools around the world,” planned for 2018.⁴¹

Outreach efforts that engage broad communities around one part of a collection may help ‘subsidize’ other, more obscure aspects or components of research collections, which are invaluable to small communities of scholarly users. This theme will come up again under “Sustainability and preservation.”

4.4.2. Unanticipated audiences

Some scholarly collections end up engaging broader public audiences than they expected to. These collections may reorient their design, development, and outreach efforts in light of

⁴¹ <http://frankenreads.org/>

unanticipated audiences. The Shelley-Godwin Archive's *Frankenstein* manuscripts launched to unexpectedly significant international public interest, particularly from Latin America and Eastern Europe (according to site analytics). International interest in collections may simply stem from the fact that they provide open access to works that cannot be found in local library collections, or in localities without libraries. Fraistat recalled the international enthusiasm for Romantic Circles, and letters he received when that collection released Wordsworth and Coleridge's *Lyrical Ballads*: "I remember we were getting emails from places in the world where somebody would thank us because they had no libraries, but they did have internet access and they could at least look at editions of the writers."

Romantic Circles continues to serve a large, and significantly public audience. According to Fraistat, the site serves about a half million unique visitors per year, over 20% of which are return visitors – "which far exceeds the number of Romanticists there are in the world, which means that people in the public are having some reason for coming back."

Unanticipated, international audiences for collections can influence how collection-creators think about the impacts of the literary texts and historical evidence they study, along with the potential impacts of their own scholarship. Fraistat expressed shock and delight at the discernable demographics of the audience for the Shelley-Godwin Archive, and suggested that awareness of these might help scholars reconsider the texts at the heart of their scholarship:

...it's a total shock in terms of understanding the contemporary reception of *Frankenstein*. Romanticist scholars – they're Western European and they're North American, mainly – and very burrowed into those horizons without most of us really understanding the lives of our texts beyond those borders –P2

Understanding broad audiences allows collection-creators to reimagine how they might create impact, engage communities, and even deploy their collection infrastructures to new kinds of purpose and contents:

At the first level we're going to be registering that worldwide, contemporary interest, and once we understand that, I think we'll be better positioned to understand how to interact with those audiences in fruitful ways, and whether our resources need to be in translations, or there are people in different countries interested in translating some of our stuff, whether we can spin off Romantic-Circles-like sites in other language communities... There's a lot to be understood.
–P2

Designing for unanticipated uses of different kinds represents a central challenge for collections as a hybrid, open-access, often experimental genre. But it is one that collection-creators have embraced, in part because they hope to broaden potential impact, but also because serving unanticipated users is imperative to the collection functioning effectively as a platform, and to the collection's ultimate sustainability. As Jewell said, "I like to be ready for the unanticipated users, or how we'll survive when I'm not around."

4.4.3. Creators as audiences

Research takes place in the construction of the resource: this is one of Palmer's (2004) central tenets, and many others have acknowledged it. As I discussed above, collection-making entails research and collaborations that serve to further scholars' work. In this sense, collection-creators themselves are an "audience" of, or benefit from, the *process* of collection-making.

In another sense, however, collection-creators also represent the main intended user groups for the actual sources that have been gathered. Speaking of the many early collections operating within the 19th-century British and American spheres, one participant said, "We talk with each other all the time. We use each other's archives" (P2). Indeed, IATH predicates its support on a collection's potential usefulness for its creator, and not for other audiences:

"[W]e say, well, how is this going to make a difference to your scholarship? And they say, well, no, those people over there are going to be really excited that we built this. And we say, no, no, what is it going to mean for you? Are you really going to be able to do something innovative in your scholarship, in your research, if we do this?" –P5

Some collections begin as an individual researcher's collection of sources for more traditional (usually monograph) publications. They are then made available online as vehicles for sharing supplementary materials, often multimedia sources that work better in digital formats than in print. In these cases, putting the collection online may mean that the creator's active use of the collected evidence for research purposes may be coming to an end, and the collection thus functions more as a supplementary publication than as a platform for further research by its creator.

Leon described her project to create a collection of digitized archival materials about Jesuit slave-ownership in 18th- and 19th-century Maryland. This collection will hold documentary evidence for an upcoming publication, funded by an NEH-Mellon Fellowship for Digital

Publication. In this way, Leon is an intended user of items of the collection, but she was quick to clarify that the collection, in its online incarnation, will not actually host the representations she uses in her own research:

The work ...will take advantage of those digitized materials, but really is working with what I call the meso-level data that has been extracted from reading the materials. ... I'll start [with census records] at 1838 and use those documents about those families at the time of sale, and work back through the archives trying to reconstruct the community. Hopefully it will get matched to digitized documents, too, but *the real work for me will be with the extracted data* –P4, emphasis added

Muñoz described how the Shelley-Godwin Archive grew in part from Neil Fraistat's own scholarly interests in textual editing, and from the difficulty of obtaining access to the fragile, original manuscripts or microfilmed versions at the Bodleian. Here again, however, the collection ultimately took a shape that was intended both to accommodate its creator's research needs, and to expand upon those to serve a wider audience. While Fraistat's (and others') editorial work might have been supported using a collection of high-quality digitized page images alone, the collection ended up as a TEI collection. The documents were encoded in part to express the scholarly insights that Fraistat and others made using page images. Muñoz noted,

...in some ways, the original image-only version [of the collection]...built for one person would've served Neil just fine. But as a way of communicating the scholarly insights about the texts that we get from the manuscripts, it seemed to me that TEI was necessary, and Neil [Fraistat, Director] agreed with that decision.

Multiple participants acknowledged using collections they built for purposes outside of research, especially for teaching. Because interview participants were the sorts of scholars who teach classes oriented toward digital methodologies, in those cases collections may not be employed for their contents as much for the technical questions they provoke:

I've used them – at least in my teaching – less for developing particular sorts of arguments and questions about a topic, and more about interrogating discovery interfaces and metadata –P1

In both of these cases in which a collection was gathered for its principal creator's own use and then re-presented as a thematic research collection, the way scholars used the collected materials was significantly different than how those materials were ultimately presented. In one

case (Shelley-Godwin), while creators made primary use of page images, extra care was put into expressing their resulting insights through encoding. In turn, these encodings enabled flexible potential uses. In the other case (Leon's collection on Jesuit slavery), digitization and access were the original purposes or imperatives of the collection. Leon's main use of the collected documents is to extract what she calls "meso-level data," and she expressed hope that the data and the resulting insights would be linked back to original documents, within the context of the collection.

Collection creators seem to get most use and benefits from their own collections in the *process* of their development and construction. Of course, that process may be ongoing indefinitely, as I discussed under "Experimentation," above. Though the immediate impacts of the collecting process may most benefit collection creators, the way collections are ultimately represented or performed on the web is usually intended to cater to broader audiences. In addition, the research outcomes of the creation process may end up being implemented as a layer of interpretation on the primary sources. Thus expressions of scholarly insight may also serve to enable new uses of the evidence. This is an interesting – and perhaps unprecedented – way for scholarly communication to build upon itself.

4.4.4. Audiences as co-creators

Many collections seek to engage audiences themselves as co-creators of the collection in different capacities. I described this briefly under "New evidence," above. Collections that do this range on a spectrum from being merely interactive, but in such a way that retains traces of interactions, to fully gathering original evidence from users of the collection.

Audiences may contribute to collections by engaging publicly, and in facilitated and documented ways, with items in the collection. Fraistat described a number of experimental efforts, in the context of the Romantic Circles collection, to integrate user responses as feedback into the collection itself. For example, the collection experimented with an informal forum-style exchange, which has been reinvigorated as a "Reviews and Receptions" space, where scholars can interact with the collection and with one another in informal ways, via book chats, reviews, forums, and more. The collection even facilitates live "Pedagogies Hangouts": "a multimedia series that brings together scholars and teachers of Romanticism at all levels to talk about the possibilities and challenges of teaching in the twenty-first century." The collection serves as a hub for scholarly interaction; and, in the process, it accumulates vibrant secondary-source content to help contextualize and enliven its collection of primary sources. Other collections, too, have encouraged

and then re-collected user interactions and social media responses. A few participants discussed this as an aspect of outreach strategies. For example, with the “Frankenreads” event mentioned above, Fraistat suggested the transformative power of collection audiences feeding their engagement back in to collection spaces as video content: “places around the world can put up video of what they’re doing, I think that will take us into new terrain.” The goal of this kind of strategy is to increase the sense of community for users of the collection. As Jewell described, part of upcoming efforts to diversify the Cather Archive and its audiences will include incorporating podcasts, social media, and other “unique content that represents that community aspect of the Cather Archive a little more, both who is creating it but also our readers and other contributing scholars and students.”

These modes of audience participation and contribution – facilitated as collection outreach through social media – do not contribute to the scholarly value of primary sources, necessarily, but they may serve to increase the contextual mass of the collection, in the sense of helping to surround items with meaningful context that can assist other users in interpretation and understanding. They also represent a variety of active scholarly communication about the collections, and exemplify a way in which collections serve as hubs for discourse.

Some collections more directly facilitate audience co-creation by supporting collaborative, distributed, volunteer transcription, encoding, enhancement, and augmentation. Fraistat explained how the Shelley-Godwin Archive planned to become a space where students and users in the public “could be meaningfully contributing to the work of the humanities.” Crowdsourcing has clear potential benefits for efficient growth and long-term public engagement with a collection. But its purposes may aim to be more universal than that. The collection aims to bolster public interest in humanities research generally, a concern that seems especially urgent in light of growing, collective anxiety about the future of federal funding for humanities research. Fraistat said, “I think initiatives like Shelley-Godwin and Romantic Circles, Frankenreads, that cross the academy with larger publics, are so important for the future of humanities” (P2).

Some collections directly solicit and collect new, original sources of evidence. The most familiar such operation is the collection of oral histories. Some collections also allow online and anonymous contributions of digitized artifacts or personal narratives. The Bracero collection, for example, “was a targeted scholarly mission to draw together a set of materials” from potential users among the public:

...the idea was that the scholars and researchers would fan out across the southwest and the California coast, the West coast, and Mexico, and do collecting days. ...they'd collaborate with local community organizations and set up a day where people would bring in their uncles and their grandparents and they'd bring all their stuff, and then they'd do digitization and oral history and then they'd submit the stuff directly through to Omeka and send people home with their collections –P4

Later in the development of the collection, project creators opened a portal for contributions directly from users of the site:

...it took a lot of convincing of the Smithsonian curators involved in the project that [unmediated collecting through a web portal] was an OK thing to do, because it was unmediated. So you'll notice if you look at the site, there's a set of contributions that have come in through the collecting portal, and they're very clearly – by design – marked as uncurated...I'm the person who decides whether something gets published or not. And the idea is that the primary standard is the material has to tell us something about the Bracero experience –P4

Collections that gather original sources from users often have to do with memorializing significant events (e.g., The September 11 Digital Archive and Hurricane Digital Memory Bank), or capturing histories for which archival documentation may be scarce (e.g., The Bracero History Archive, which “collects and makes available the oral histories and artifacts pertaining to the Bracero program, a guest worker initiative that spanned the years 1942-1964”).

The priority that public-oriented collections give to scholarship varies. The principal distinction between these and other thematic research collections is the extent to which they are curated. The September 11 Digital Archive originated in the Public Projects division of the Roy Rosenzweig Center for History and New Media. Scholars and researchers created the collection, and members of the public filled it up with contributions of text stories, photos, videos, etc. The parameters for contribution were liberal; the filtering and assessment of authenticity were minimal. The collection is not curated in the sense that most research collections are. But the infrastructure was generated by historians, and it remains the most comprehensive collection of historical evidence about the lived experiences of people on September 11, 2001. Indeed, it became the Library of Congress' first major digital acquisition in 2003.

Of course, ethical collection development is always a concern, and this concern is compounded when users have opportunities to be highly interactive with collections. For collections built around vulnerable populations or communities, the design of the collection as a

platform must take into account the values and epistemologies of those communities. This is not a simple proposition. One participant asked, of creating interactive collections, “How do you make it possible? And how do you make it more than possible? How do you make it attractive? How do you make it fun? How do you make it safe, in some cases, depending on the community that you’re working with? How do you make it respectful for people to be able to do new things with a collection?” (P7).

For example, for collections that are built for purposes of social and political activism – including for history and research thereof – interactivity and participation may entail risk for users. In those cases, participants acknowledge the imperative that collection-creators, in addition to technologists and tools that facilitate collection-building, consider account user privacy in the design of interactivity and contribution mechanisms. For example, Leon pointed out that most or all of the collecting projects around the Black Lives Matter movement published in the last few years were built using Omeka and its Contribution plugin: “We discovered from that work that there was a baseline setting in that suite that the project directors were pretty sure put people in danger – and that was that they could not contribute completely anonymously.” Fixing this problem became an immediate priority for collection creators and, concomitantly, Omeka developers.

The multiplicity of purposes and audiences has obvious benefits for thematic research collections as a genre, but it also has drawbacks. Some have suggested that the hybridity of the genre has problematized its valorization in systems of scholarly evaluation (Thomas, 2016, among others). Another participant suggested that thematic research collections, in trying to appeal to everyone, may end up appealing to no one in particular, in part for lack of recognizable generic features:

...they don’t have the same sense of belonging to a genre that you do with a documentary edition, where the form of the published object has features that you expect because you’re familiar with the practice of documentary editing or the genre of documentary edition. ... they don’t arise out of a tradition, or they arise out of multiple traditions, so one of the things that shapes each of those idiosyncratic objects is the history of scholarship on that particular subject... which is one reason why often these things look and feel like something that was designed for experts in the subject matter, and not for a general reader, and they don’t belong to a category of things that’s familiar. You could suppose that if this went on for a hundred years or so, that they’d develop some conventions, but right

now they differ pretty much one to another, which is another reason that librarians hate them –P9

Muñoz noted that perceiving the Shelley-Godwin Archive as blurring boundaries between research, teaching, and service was not so much an intentional aspect of design as a prism for understanding the ultimate contributions of the collection:

...the way in which it might be a crossover between research, teaching, and service is something that's easier to see in retrospect...I don't know how much of that was explicit at the time. But it's certainly a useful prism in looking back.

Specifically, understanding that collection as a hybrid between teaching and research enables the collection's creators to perceive shifts in the collection's value over time:

...the research part maybe was biggest at the beginning, and is kind of tailing off, whereas the teaching part was smallest at the beginning, [but] it's growing... in some ways I think of its pedagogical usefulness as core to the argument for its long term stewardship, more than the research" (P6)

In summary, this research demonstrates the collections are built around a multiplicity of purposes, and aim to serve a corresponding diversity of audiences. Upon the foundational purposes of collocation, access-provision, and some level of interpretation and activity support, collections aim to be generative of experimentation, collaboration, new research directions, and new evidence. Chapter 7 returns to the notion of generativity by considering how collections serve as platforms for research. I have devoted a chapter to this exposition of collection purposes because purpose is a defining characteristic of collections, and it helps determine and shape other defining features, including – I argue – different broad *kinds* of collections.

CHAPTER 5: KINDS OF COLLECTIONS

5.1. INTRODUCTION

This chapter returns to the question of the defining features of thematic research collections as a scholarly genre, particularly what *kinds* of thematic research collections there are, and how they may be distinguished from one another. This study aimed to comprehend the breadth and variety of the landscape of collections through typological analysis. I aimed to enrich the outcomes of typology with content analysis, which provides deeper insight into collections selected as exemplars of preliminary types. Together, these approaches suggested three useful kinds of collection to distinguish:

- Definitive-source collections (Type 1)
- Exemplar/context collections (Type 2)
- Evidential platform collections (Type 3)

The main distinguish property of these types is their central purpose, or their main, intended contribution to scholarly work. Each type gathers items in pursuit of a different kind of completeness. I will discuss how a collection's purposes suggest or help determine the completeness that it is built toward. In turn, a number of other context- and content-related attributes differ between these types. I will return to a full exposition at the end of this chapter, stepping through the attributes to paint a picture of these types as representing three prominent (but not exclusive) ways in which collections both manifest and contribute to research.

To identify different types may seem like a presumptuous undertaking when our definition of “thematic research collection” as a whole, in deference to the flexibility of digital scholarship, still draws a loose and changeable line around the objects of study. Ontologically speaking, there are probably no types (or, speaking colloquially, about as many types as there are thematic research collections). But by identifying common constellations of properties, we create a navigational tool: a common sense of the diversity of the genre, a shared vocabulary about the choices that collection-creators make, the various things collections aim to do, and significant and meaningful variations in their architectures and affordances.

This chapter first discusses the outcomes of the provisional, formal typological analysis. I then describe how findings of the content analysis clarified and enriched that typological effort to produce the final three types presented above. I define several attributes that help distinguish

collections, with a focus on the concept of *completeness*. The final section of this chapter describes the resulting typology, and makes the case for its usefulness as a frame for understanding the genre.

5.2. PROVISIONAL TYPOLOGY⁴²

There are numerous conceivable ways to approach the division of a mass of collections into types. As discussed in Chapter 3, I began by examining the sample of approximately 100 collections for potentially discriminating attributes. I was especially interested in the data models underlying the collections, in the first run at typological division, because data models are suggestive of and help determine other meaningful attributes: How do collections facilitate access to items? How do collections represent items in relation to one another, and to contextual information and other resources?

The first approach at typological analysis revealed something surprising: not all collections provide straightforward access to their items. In other words, not all collections provide search and browse across the primary sources that they gather. This is counterintuitive; search and browse are implicit in our account of what thematic research collections basically do, which is gather and *provide access*. I found that some collections mediate access to items in important ways, sometimes through external affordances that manifest scholarly interpretation, and sometimes by foregrounding pieces or derivatives of items over facsimile or other (attempts at) complete representations. This was a first clue to how collections perform their purposes differently.

The provisional typology relied upon two properties in particular of the conceptual data models of collections:

- The first property asks what priority the collection's model gives to primary sources, in terms of their visibility and accessibility in the collection relative to other scholarly content? In short, are primary sources the main content of the collection or are they ancillary? This property is usually reflected in how a collection is navigated and how search results are presented.

⁴² This chapter borrows from Fenlon (2017).

- The second property concerns whether the collection employs advanced markup, which may be considered deeply descriptive, and which enables functionality beyond basic keyword searches.

Though there are myriad properties that may be used to describe the data models underlying collections, these two serve to enable and constrain the use of collections in fundamental ways; they play significant roles in determining how collections are developed, what subsequent data models are employed, the uses to which collections can be put, and much more. Table 6 provides more detail on these two properties, which are treated as binary properties for the sake of typological distinction, but which in reality are probably more readily recognized as being on a spectrum.

Table 6. Provisional properties of collection types

Direct item-level access to primary sources <ul style="list-style-type: none"> • It is clear what the primary sources are • Primary sources constitute main contribution • Primary sources directly accessible through search and browse 	Indirect or mediated access to primary sources <ul style="list-style-type: none"> • Search and browse do not operate directly upon primary sources • Access to and visibility of primary sources is mediated by an analytic or interpretive layer • Analytic or interpretive layer constitutes main contribution
Advanced markup <ul style="list-style-type: none"> • Items are encoded with rich markup that supports advanced and multimodal representation 	Minimal markup <ul style="list-style-type: none"> • Items are encoded minimally, to the extent they must be for presentation on the web or to enable keyword search across texts.

In this first phase of typological analysis, I isolated a third attribute of collections, which also helps determine forms and functions of collections, albeit from a different angle – whether a collection seemed to be oriented primarily toward supporting research, supporting teaching, or gathering new, original evidence. Because supporting research is the predominant goal of collections, this type was by far the most common. In later rounds of analysis, my view of collection purpose became both more refined and more central to the typological account.

To give a sense of how well each type was represented in the sample, Figure 8 gives the number of collections originally deemed to present each combination of properties among the subset of collections with the primary purpose of providing research support. A further 8

collections were deemed to primarily serve the purpose of teaching (Provisional-type 4), and 19 to serve the purpose of soliciting new evidence (Provisional-type 5). Provisional-type 2 collections were most common in the initial sample, followed by Provisional-type 1 collections. Note that the sample was eventually expanded, and typological boundaries shifted, as discussed in the next section.

	(A) Direct access	(B) Indirect access
(C) Advanced markup	26	1
(D) Minimal markup	34	9

↪ Type 1
 ↪ Type 2
 ↪ Type 3

Figure 8. Number of collections originally falling into provisional types 1-3

Given these three properties of (1) access to primary sources, (2) markup provisions, and (3) main purpose, the provisional typology could be imagined in a three-dimensional property space (Figure 9).

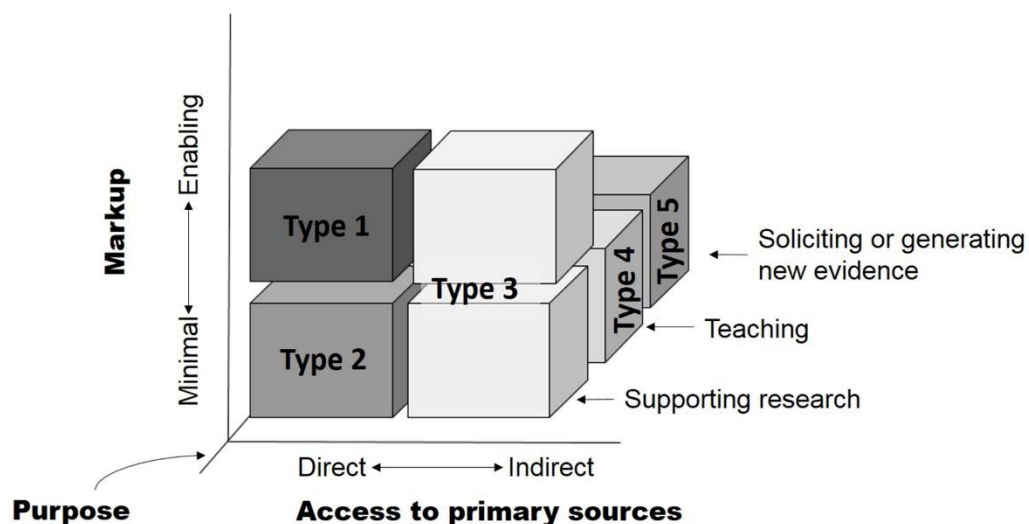


Figure 9. Provisional division of types, shown in 3-dimensional property space

The provisional typology identified the following five kinds of collections:

Provisional-type 1 collections provide direct access to primary sources along with advanced markup, which is markup that enables access to texts beyond rendering them and affording keyword searches. Encoded primary sources – predominantly texts – constitute the main content of these collections, although many include extensive image content and other sources that are devoid of markup. Secondary sources and other kinds of information contextualize and supplement the primary-source content. To literary scholars these should be familiar types of collections; among them are the most well-known, oft-cited, and longest-running thematic research collections, including Thomas MacGreevy Archive, Walt Whitman Archive, and World of Dante. Provisional-type 1 collections are often, although not always, self-described as archives and aim to be comprehensive authorities on the literary works of a particular author, or group of authors circumscribed by time period, proximity, or social relationship.

Provisional-type 2 collections also provide direct access to primary sources, but these collections afford minimal markup for various reasons. Many of these collections gather heterogeneous media and formats. When gathering text they tend to place less emphasis on affording fine-grained access to texts (as in provisional-type 1 collections) and more on providing other kinds of research support, such as comparative views of high-quality images or embedding primary sources within scholarly narratives.

Provisional-type 3 collections provide indirect or mediated access to primary sources. These may be understood as data-centric or derivative-centric collections, and they are of interest because they stand at the very edge of our usual understanding of collections. While these collections include primary sources, they are not always directly accessible as such, or they are not prioritized in the navigation and design of the collection. Rather, access to primary sources may be mediated by an analytic or interpretive layer, such as an interactive map or 3D model, or else the collection may afford access to more granular derivatives of primary sources. In this way, the collection primarily offers or makes most visible data gleaned from primary sources. Exemplars of provisional-type 3 include *Voting Viva Voce: Unlocking the Social Logic of Past Politics*, and *Aquae Urbis Romae: The Waters of the City of Rome*.

Provisional-type 4 collections make teaching (as opposed to research) a central focus, or are built for pedagogical purposes. Collections may cater to any level of education or the general public. They may provide either direct or indirect access to primary sources, and they may use markup or not (although all of the collections surveyed employed minimal markup). The rationale

for distinguishing pedagogical collections as a separate type, despite the particularities of their data models, is that this study is concerned with collections in practice. The distinctive purposes and audiences determine how such collections are developed, evaluated, used, and managed. Examples of provisional-type 4 collections include Salisbury Project and I Am a Man: The Memphis Sanitation Workers' Strike.

Provisional-type 5 collections are primarily intended to solicit or generate new evidence, including oral histories and digitized artifacts. New sources are simultaneously created and collected. This sets them apart from collections that curate or aggregate existing archival, literary and historical sources, often from extant collections. These collections may provide direct or indirect access to primary sources, but, as with provisional-type 4 collections, they tend to manifest only minimal markup. Collections that solicit original evidence may thus resemble one of the above types, but they are distinguished by the scope and processes of their development, which differ fundamentally from those of other collections. Examples include Voices of the Jazz Era Ballroom and the Rabat Genizah Project.

I do not go further into an explication of these types or their implications for our understanding of the genre, in favor of turning to the refined types that emerged through content analysis. A fuller account of the provisional typology is available in Fenlon (2017). This formal typological effort yielded mutually exclusive categories of collections which are sufficiently suggestive of potentially useful types. However, the resulting types are not intuitively recognizable. Types are meant to serve as handles for grasping and pulling at complex phenomena. In that role, these provisional types offer a less than satisfying grip on the bulk of thematic research collections.

However, the provisional typology effectively serves to ground and inform the next stage of inquiry, an in-depth content analysis of representative collections, as described in "Methods."

5.3. ENRICHING TYPOLOGY WITH CONTENT AND CONTEXT

Chapter 4 examined the purposes of collections as a defining feature of the genre. Purpose emerged as a significant property from the content analysis of collections and serves to shape other aspects of collections. After I completed the provisional typology, I selected collections to analyze as exemplars of each of the three central types, as described in section 3.3. The process of developing the detailed analysis protocol for qualitative content analysis refined my sense of

collections' attributes and how they interrelate. In particular, content analysis suggested strong linkages between the defining purposes of a collection other aspects of its contents – its theme, the diversity of its items, etc. In the course of representative collections in depth, and in terms of their purposes, I found that the original typological properties were crude proxies for the purposes of collections. The second stage of typological analysis was refined by this outcome of the content analysis, and focused on how the forms of collections can be traced to their originating purposes.

Each of the three collections subjected to qualitative content analysis represents one of the three main, provisional types, as described in Chapter 3:

- Provisional-type 1: The Shelley-Godwin Archive
- Provisional-type 2: The Vault at Pfaff's
- Provisional-type 3: O Say Can You See: Early Washington, D.C., Law & Family

Table 7 revisits the overview of attributes employed in the content analysis, described at length in Chapter 3 and detailed in Appendix B.

Table 7. Overview of content analysis protocol

Cluster	Categories of analysis
Context	Theme; Purposes; Impact; Creators; Audience; Documentation; Provenance; Related collections; Related projects and publications; Review; Funding; Developmental stage; Host; Rights; Sustainability and preservation plans; Method
Content	Items; Diversity; Size; Narrativity; Quality; Language; Completeness; Density; Spatial coverage; Temporal coverage; Interrelatedness
Design	Data models; Navigation; Infrastructural components; Interface design; Interactivity; Interoperability; Openness; Identification and citation; Modes of access and acquisition; Accessibility; Flexibility

The refined typology draws on the following attributes of collections:

- From the Context cluster: purpose, theme
- From the Content cluster: items, interrelatedness, diversity, completeness

Though data models were central to provisional types, the Design cluster does not play into the refined typology directly because, through the analysis of collections in their fuller contexts, I

found it less useful to locate typological distinctions in technological designs. Rather, I see the design choices, ranging from data model selection to interactive affordances, as *implications of* typological distinctions described here, which are themselves determined foremost by collection purposes.

Of all of the attributes identified by the content-analysis protocol, these six were identified to scope the typology because they have the most visible implications for collection purposes and how they are manifested in the forms of collections. The strength of these attributes is that in combination with one another they distinguish different kinds – among collections with similar *purposes*, we witness similar kinds of *themes*, similar ways of representing *items*, similar kinds of *interrelationships* between items, and the pursuit of the same ideals of *completeness*. Other attributes identified in the content analysis show differences among themselves, but those differences seem less to pertain to differences in *kind*, and more to differences in implementation or realization. Therefore, the attributes selected for description here are the main and most interesting distinguishing attributes.

Below I define each of these select attributes, and demonstrate how they play into each proposed type. For more detail on each of these attributes, see Appendix B. “Completeness” is afforded its own section (5.4, below), because it stands out – in typology, content analysis, and interviews – as essential and especially revelatory of differences among collections.

5.3.1. Theme

A collection’s theme is its subject or topic, its “controlling idea” (Kuhn, Johnson and Lopez, 2010) or “conceptual core” (Mattern, 2012). A theme can be defined around an author, work, event, phenomenon, or any object of study (Palmer, 2004). It may be broadly or narrowly defined. “Theme” is related to the concept of unity criteria, “the criteria that determine whether an item is gathered into a particular collection...a formulation of the decision-making process that guides the development of a collection and captures the curator’s intent” (Doerr, 2014; Wickett, et al., 2013). However, precise selection criteria may be driven by more factors than just theme. In this scheme of content analysis, unity criteria may instead be understood as triangulating and operationalizing a few different attributes from this protocol: theme, completeness, items, diversity, perhaps more. Martin had an interesting take on the notion of *theme*, not just an abstraction, but as made manifest in the “connective tissue” of a collection:

...when you say “collection” or even “repository” or “archive,” one tends to think of just the materials. So you just think of something that might look like a directory listing of a whole bunch of files of page images of the primary documents, and leave off all the connective tissue, all the different kinds of indexing forms, and other kinds of things that ...might be called metadata by some people, but it’s more than what would generally be thought of as metadata – much more. That’s the thematic part, right?

5.3.2. Items

This attribute asks, what constitutes an item in this collection? We have considered an item, simply, as something that has been gathered into a collection. Content analysis reveals that collections are constituted of many kinds of things. Often the conceptual unit that a user might identify as an item of interest may not be located in any particular digital object, but in the interrelation of many different digital objects and processes. There are distinctions between the things a collection purports to gather, the things a collection actually makes accessible as such, and the layers of representation that underlie those things with which users interact. Which of these is a collection “gathering”? It is clear that “item” must be defined in part by context. For example, what a user sees as an item, a curator may see as a component of a more holistic piece, and a developer may see as a compound-digital-object-plus-associated-processing-routines.

This category considers how we can characterize discrete items in the collection, holding that term loosely, and focusing on the conceptual. One clue lies in what units are returned by searching and browsing mechanisms. This attribute takes as a condition the fact that, by our definition of *thematic research collection*, primary sources are usually the main “items” within these collections. But there are complications here, as we shall see. It remains an open question to what extent the “connective tissue” or interpretive layers and affordances can be considered “items” of a collection. In this regard, collections seem to fall along a spectrum; in some extreme cases, primary sources are present but so tightly interwoven with and mediated by interpretive affordances that it is not clear where the boundaries of “items” lie. Flanders (2014) invites us to “consider what happens to our understanding of a ‘collection’ when its constituent items are no longer the primary unit of meaning”; so this analysis does.

5.3.3. Interrelatedness

Beyond the “is gathered into” relation, which is the defining relationship of collections that obtains between collections and items (Renear, et al., 2008), what other relationships occur between items (and other things) in a collection? How do those relationships help constitute

thematic research collections? This attribute is not necessarily concerned with logical relationships in this phase of analysis, but with informal characterization of relationship types and how they are implemented. Palmer (2004) refers to thematic research collections as “dense, interrelated” collections, suggesting that interrelatedness is a central component of contextual mass. But we have yet to determine how collections interrelate items, or illuminate relationships between items, in order to enact their goals. Of course, one immediate relationship is forged by collection itself, as the act of gathering, of bringing related things into the same space. Collection, in this case, is the assertion of one kind of relatedness: relevance to a common theme. Flanders (2014) considers a similar question: “how do the boundedness and *internal cohesion* of a collection help to define its intellectual purpose?” (Emphasis added). In this analysis, I consider the question from the opposite direction: how is purpose reflected by the internal cohesion of a collection, which is determined (in part) by interrelationships among a collection’s items? This category will serve to exemplify some of the ways in which items are related to one another and to other things within and beyond collections.

5.3.4. Diversity

The notion that collections are diverse in their contents is central to the earliest accounts of thematic research collections. Palmer (2004) balances the thematic *coherence* of collections against their heterogeneity. Diversity of sources, by Palmer’s account, is key to creating platforms that can serve users across disciplinary boundaries: “the aggregation of diverse sources – images, texts, numerical data, maps, and models – will seed intellectual interaction [among diverse intellectual communities] by making it possible to discover new visual, textual, and statistical relationships within the collection and between lines of research” (Palmer, 2004). Many of the sources reviewed from digital humanities literature suggested that diversity in the form of multimedia content should be a priority of digital collections. This is one expression of a very common theme encountered in the best-practices and evaluation literature, which is that digital scholarship must attempt to exceed the capabilities of print scholarship: “Coherence, then, refers to the graceful balance of familiar scholarly gestures and multimedia expression which mobilizes the scholarship in new ways” (Mattern, 2012). A second perceived benefit of item diversity is that it enables “corroboration” among forms of evidence: “Does it effectively ‘triangulate’ a variety of sources and make use of a variety of media formats?” (Mattern, 2012). Diversity has long been a

value ascribed to historical evidence. In his highly cited study of historical methods, Bloch (1954) asserts:

It would be sheer fantasy to imagine that for each historical problem there is a unique type of document with a specific sort of use. On the contrary, the deeper the research, the more the light of the evidence must converge from sources of many different kinds. –Bloch, 1954

We will revisit each of these attributes, using examples from the content analysis, to show how each appears in and contributes to the refined typology described below.

5.4. COMPLETENESS

The accounts of provisional types (above) mention the notion of “completeness” in passing, observing that provisional-type 1 collections “aim to be *comprehensive authorities*...”. The notion of completeness becomes more central to this account of types in light of an in-depth study of collection purposes, through content analysis and interviews.

Collections strive toward different ideals of completeness, closely related to their overarching purposes and the scholarly traditions of their creators. Muñoz articulated how a certain “logic of completeness” motivated the Shelley-Godwin Archive and its collection-development strategies. The Shelley-Godwin Archive aims to collect all the known, digitized manuscripts of Percy Bysshe Shelley, Mary Wollstonecraft Shelley, William Godwin, and Mary Wollstonecraft – “England’s first family of writers.” But by Muñoz’s account, the collection originated with Neil Fraistat’s interests in Percy Shelley’s manuscripts alone (Fraistat being a Percy Shelley scholar, editor of *The Complete Poetry of Percy Bysshe Shelley* and the Norton Critical edition, *Shelley’s Poetry and Prose*). Mary Shelley, author of *Frankenstein*, was included in part to broaden the collection’s potential audiences. Muñoz calls Mary Shelley “charismatic megafauna”; her manuscripts were included in part to broaden the collection’s appeal, which will help to sustain it on behalf of the more obscure works of niche scholarly interest. And because “grant-getting rhetoric thrives on completeness” (P6), collection creators thus expanded the scope of the collection to encompass a more “charismatic” or broadly appealing notion of completeness:

So, Percy Shelley and Mary Shelley? Well, why not Mary Shelley’s mother, also famous, and William Godwin, her father, also famous? And if you have all four of them, then you have the first family of British Romanticism, and that is itself a kind of charismatic megafauna. –P6

Of course, certain practical considerations came into play. For example, the collection creators had preexisting relationships with the institutions that hold the largest concentration of relevant, original sources. Around ninety percent of the manuscripts live at Oxford's Bodleian Libraries, with which MITH had an established, successful collaboration on the Shakespeare Quartos Archive.

Nevertheless, the completeness toward which the Shelley-Godwin Archive bends is the same one that guides scholarly editions, and this is not coincidental:

...the 'edited complete works' is a genre in textual studies, and so [the Shelley-Godwin Archive] also had its own logic of completeness because of Neil [Fraistat's] disciplinary home, though I don't know that that was ever explicit. But he was about completeness in his analog humanist life, so there was a logic to it in his digital life. ... the way that we did it relies mostly on the canonical sense of what [completeness] means from the traditions that we've received –P6

In the case of this collection, then, the completeness that drives collection development stems in part from the generic tradition of scholarly editing. Unsworth also pointed to the notion that predecessor genres of scholarship can bound and shape collections, using the Rossetti Archive as an example:

...they arise out of multiple traditions, so one of the things that shapes each of those idiosyncratic objects is the history of scholarship on that particular subject, and what are the catalogs or the indexes that people use as standard reference points to develop a common nomenclature for objects in the universe of stuff that Rossetti produced. The Surtees Catalog is Rossetti's.

Sources used to create the content analysis protocol (which clued me in to the notion of "completeness" as an attribute of collections), asked questions like, "Is relevant content missing? Is any omission explained and/or justified?" (Sahle, et al., 2014); whether the collection makes available "what one would expect such a resource to provide" (Mandell, 2012); and whether a collection "claims to be representative for a specific subject domain" or "functions as a reference for that domain" (Henny, et al., 2017). Here we find clues that completeness is integral to purpose: what does a collection intend to contribute to scholarship, and therefore what must it seek out to collect? One resource considers completeness to be a principal of collection design: "Which principles guide the design of the text collection, does it for example aim at completeness, representativeness, balance, exemplarity?" (Sahle, et al., 2014).

While many collections aim to be exhaustive and authoritative on a particular subject, or within a particular authorial oeuvre, other collections are satisfied by other kinds of completeness. For example, what would it mean for a collection about “Nineteenth Century Disability: Cultures and Contexts” to be complete? Would creators need to pursue every extant, possibly relevant artifact, or every primary source text from the 19th century that mentions disability? This seems infeasible for a single collecting project. What about “The Countryside Transformed,” a collection on the railroad and the eastern shore of Virginia around the turn of the century, which aims to explore the “physical and mental landscapes in which the people of the region lived, worked, and traveled⁴³? The notion of completeness at the hearts of these collections is certainly not as straightforward as that which motivates the Shelley-Godwin Archive, or the Walt Whitman Archive, or the William Blake Archive.

We might find a clue toward another kind of completeness in the concept of contextual mass (though that principal can be witnessed in all thematic research collections), which calls for “dense, interrelated” collection that provide a supportive context for research. These collections support research not by being exhaustive, which may be impossible given the theme that unifies the collection, but by gathering exemplars and rich contextual information. They seek out *exemplarity*, and to surround representative primary sources with rich contextual information, including secondary sources.

In the content analysis of the Vault at Pfaff’s we find an example of this kind of orientation. Vault at Pfaff’s aims to “gather and organize both primary and secondary source documents about the bohemians of antebellum New York.”⁴⁴ This is a vast landscape with indistinct borders. Collecting exhaustively according to these criteria would be infeasible; clearly articulating an exhaustive completeness for this theme would be impossible. Instead, Vault at Pfaff’s explicitly employs loose, inclusive selection criteria, offloading the ascertainment of authenticity and ultimate relevance to its users, each according to their own interests: the collection aims to “allow students of American culture to determine who the Pfaff’s bohemians really were and to assess their contributions to the art and literature of the antebellum United States.”⁴⁵

⁴³ <http://eshore.iath.virginia.edu/>

⁴⁴ <https://pfaffs.web.lehigh.edu/node/38090>

⁴⁵ <https://pfaffs.web.lehigh.edu/node/38090>

The content analysis of provisional-type 3 collection O Say Can You See: Early Washington, D.C., Law & Family suggests a third kind of completeness. O Say is motivated by a different sort of purpose than the other two analyzed collections (as discussed in Chapter 4). It aims toward a more specific research objective: to analyze “multigenerational black, white, and mixed family networks in early Washington, D.C.,” using “case files from the Circuit Court for the District of Columbia, Maryland state courts, and the U.S. Supreme Court.”⁴⁶ In other words, this collection is more explicitly interpretive than the others; as such, its completeness revolves around the sufficiency of its evidence. Are the sources that this collection gathers sufficient to satisfy the analytic goals of the collection? If so, the collection is probably complete. Indeed, as this collection grows, it may grow toward new analytic goals, or the confirmation of extant analytic outcomes using new sources. In other words, this collection pursues completeness, not in the sense of being exhaustive or exemplary, but in the sense of being evidentially sufficient. This is not the same as evidential necessity; the collection does not limit itself to strictly necessary sources. Beyond the legal documents – petitions for freedom, and civil, criminal, and chancery case files from different archives – which constitute their main sources of evidence, the collection gathers related documents wherever possible from churches, archives, special collections, and historical societies.

Content analysis and interviews therefore suggest three distinct notions of completeness for thematic research collections:

- *Definitiveness*: sources are the “most authoritative of their kind” (OED, 2017), and therefore unique, high-quality, and authentic. The collection successfully guided by this kind of completeness may be considered exhaustive, or, as Unsworth put it, “definitive with respect to purpose.”
- *Exemplarity*: often the guiding light in cases in which definitiveness or exhaustiveness is impossible or infeasible, the sources aim to be representative and exemplary. Completeness criteria are inexact, and vary with respect to theme.
- *Evidential sufficiency*: sources are gathered to provide sufficient evidence for a specific interpretive or analytic goal.

⁴⁶ <http://earlywashingtondc.org/about>

The different varieties of completeness align with and complement collection purposes to discriminate different types of collections; as such, they are reflected in the names of the types described in the next section.

5.5. PROPOSED KINDS OF COLLECTIONS

When we look across the landscape of thematic research collections, we can distinguish three apparent pools of collections. These types are centered on three different, overarching missions, and guided in development by three different kinds of completeness:

- **Definitive-source collections (Type 1):** The central and distinguishing objective of these collections is to bring together *definitive* literary or historical sources, upon which layers of affordance and interpretation may be built to varying degrees. These collections are driven by qualities of their items, the sources they gather. These collections aim to reassemble the human record in digital form (Thomas, 2016), and shape its affordances.
- **Exemplar/context collections (Type 2):** These collections aim to gather a dense base of *exemplary* items (with respect to their themes), surround those items with contextual information, and illuminate relationships among them. These collections have a principal objective of discovering and illuminating relationships between the items they gather, and clearly suggest the principal of contextual mass (Palmer, 2004).
- **Evidential platform collections (Type 3):** These collections gather primary sources, but treat them differently, in part because they are driven by specific interpretive or analytic goals. The main goal is not to make primary sources accessible as such (though they almost always are), but to interpret and leverage kinds of *evidence* derived from these sources into flexible platforms for new kinds of interpretation. These collections aim to identify and realize the evidential potential of their sources, to be extracted and manipulated into serving as evidence for a specific research objective.

These types are based on provisional-types 1, 2, and 3, which emerged from formal typological analysis, and as such their definitions bear clear resemblance to those provisional descriptions. However, they are refined by close analysis of collection purposes and related attributes. These types focus on purposes rather than specific aspects of data models. Table 8 gives

the number of collections that fall into each refined type, mapped from the original, provisional types.

Table 8. Number of collections in each type, mapped from provisional types

<i>Provisional type</i>		<i>Type 1</i>	<i>Type 2</i>	<i>Type 3</i>
<i>1</i>	mapped to...	25		
<i>2</i>	mapped to...	11	23	
<i>3</i>	mapped to...	1	2	8
<i>4</i>	mapped to...		8	
<i>5</i>	mapped to...	2	16	1
<i>Additional sample</i>		9	33	6
<i>Total</i>		48	82	15

Table 9 details the three kinds of collection. Each column represents a type. The “Examples” row gives a few examples of collections that fall into each type. (I have included a very brief statement of the topics of Type 3 examples, because their titles are less intuitive). The “Principal purpose” row briefly states the main, intended contributions of each collection (reducing the purposes discussed in Chapter 4). Purpose is the main basis of division in this typology, from which all others extend. The remaining rows correspond to attributes defined above. In the next section I elaborate each type, stepping through the attributes given in the table, with more explanation and examples. I also discuss what kinds of use each type may enable.

Type	(1) Definitive-source collection	(2) Exemplar/context collection	(3) Evidential platform collection
Brief definition	Bring together definitive sources and add affordances: “ <i>reassembling the human record in digital form</i> ” (Thomas, 2016), and shaping its affordances	Interrelate and (re-) contextualize diverse sources: “ <i>creating dense, interrelated collections</i> ” (Palmer, 2004), by building context within and around materials	Aggregate, deconstruct, remodel sources for new uses: “ <i>a negotiation of meaning</i> ” (Flanders, 2015), leveraging evidence into flexible platforms
Themes	An author, group of authors (bound by period or place), or work	A concept, event, place, or phenomena in between	A specific research question or objective
Examples	<ul style="list-style-type: none"> • <i>The William Blake Archive</i> • <i>World of Dante</i> • <i>Romantic Circles</i> • <i>Journals of Lewis and Clark</i> 	<ul style="list-style-type: none"> • <i>Salem Witch Trials</i> • <i>Uncle Tom’s Cabin and American Culture</i> • <i>Black Gotham Archive</i> • <i>Nineteenth-Century Disability: Cultures and Contexts</i> 	<ul style="list-style-type: none"> • <i>Valley of the Shadow</i> – life in two American communities during the Civil War era; • <i>Voting Viva Voce</i> – networks that underpinned political activity in the era of voting by voice
Principal purpose	<i>Collocate</i> and provide <i>advanced access</i> to <i>high-quality, unique, and value-added</i> representations of sources	<i>Gather</i> and imbue <i>diverse sources</i> with <i>rich, interpretive context</i> and <i>actionable interrelationships</i> between components	<i>Refine, integrate, and arrange diverse sources and derivatives</i> into <i>platforms</i> for new kinds of <i>analysis and interpretation</i>
Items	Primary sources or digital editions as central, <i>conceptual</i> unit of gathering; other materials distinct, supplemental	Primary sources <i>may</i> be conceptual unit, with discovery and access mediated by an interpretive layer; metadata as main unit of gathering, discovery, and access	Gathers primary sources to anchor the conceptual, abstracted entities or derived data, interpretation, and analysis at the forefront
Completeness	Definitive	Exemplary	Evidentially sufficient
Interrelatedness	Forges relationships between items based on literary and bibliographic properties, using markup exploited by advanced reading and comparative facilities	Illuminates historical and literary relationships between items, entities, and components, using metadata, interpretive discovery layers, and hyperlinked narratives	Interrelates sources by extracting and integrating data; relates data to surrounding documents and context through narrative, navigation, and interpretive layers
Diversity	Not necessarily of primary sources, but through incorporation of often extensive, diverse secondary source material	Diverse kinds of sources, on diverse topics, often interwoven with secondary sources and other interpretive affordances: interdisciplinary platforms [1].	Integrating discrete, homogeneous sets of data derived from different sources, corroborated with diverse other forms of evidence and interpretation.

Table 9. Overview of types

5.5.1. Definitive-source collections

The distinguishing, main objective of these collections is to bring together *definitive* literary or historical sources, upon which layers of affordance and interpretation may be built to varying degrees.

By “definitive,” I do mean that they gather exceptionally high quality and potentially unique primary sources, sometimes in the shape of digital scholarly editions of literary works. Even when they are not digital editions per se, the same level of editorial attention may go into their representation: their digitization, transcription, encoding, and visual representation. These collections are driven by qualities of their items, or the sources they gather. This is a familiar form to the digital humanities community. Interview participants referred to collections of this kind as “TEI projects” (though a few collections that fall into other types employ TEI as well). Many of these collections may emerge from the tradition of scholarly editing (as discussed in section 5.4, above), and can be seen to share and expand upon its goals.

By “bring together,” I mean that often what makes these collections definitive is not only the quality of their materials; it is the fact that they unify access to distributed materials of high quality. The digitized primary sources may exist elsewhere, especially in the digital collections of original holding institutions, but definitive-source collections serve to unify access to them along the new collecting logic of their “theme.” As Palmer (2004) notes of thematic collections generally, “The physical proximity of resources becomes trivial when the material is digital and made available in a networked information system (Lagoze and Fielding, 1998), but the intellectual and technical work of selecting and structuring meaningful groupings of materials remains critical.”

Jewell described the imperative for the Willa Cather Archive: “more and more I wanted it to be unique things that aren’t available elsewhere. We’ll still provide digital versions of texts just for convenience, but the real focus of our energies is on unique contributions.” In this statement, Jewell contrasts the provision of texts for *convenience* with the provision of *unique* content, suggesting that there’s a difference in kind of intended contribution. Uniqueness appears to be related to the notion of definitiveness, if only because it is implicit in the combined attributes of *exhaustive* and *authoritative*.

However, there is an aspect of convenience driving some definitive-source collections, particularly if we take “convenience” to be a motive for collocation. A collection may gather

unique items. A collection may also gather items uniquely, in the sense that no other such gathering exists to support the same research purposes. The gathering itself may be definitive, exhaustive, even authoritative, with respect to the theme of the collection. For example, the Shelley-Godwin archive gathers around a group of authors. Not only are the digital manuscript page images authoritative (in the sense of being high-quality images created by the original, hosting institutions); and not only are the encodings definitive (in the sense of being scholar-generated encodings validated against a well-documented editorial scheme) but the collocation itself is unique. The collection unifies items not only by author, which is already uniting dispersed resources, but by an interrelated group of authors, so that there are more, potential connections and interpretive insights waiting in the collection.

Of course, collection development takes a significant amount of time. In the pursuit of definitiveness, a collection may take up and drop threads of development, shifting course in response to audience, changes in the copyright status of works, or available funding. Jewell describes how the Cather Archive began by pursuing juvenilia, as unique, unpublished Cather content, but transitioned to Cather's more popular letters when copyright restrictions dissolved:

...there's always an interest in trying to provide unique content on the Cather Archive, but that isn't easy, especially if you're interested in unique textual content for the study of an author...Our initial impulses were to go to some of her early journalism...[but] we didn't have a lot of scholars that interested in that early, juvenilia stuff. *But it was what hadn't been published before...*

...when [publication of the letters] became available, it was clearly and easily the priority within the world of Cather studies to focus on making that material available for everybody....I feel that in some ways this edition we're working on now is what I always wanted the Cather Archive to be, but it hasn't quite been before.

Definitive-source collections are not always diverse in terms of the media or types of primary sources they provide, but may be diverse in their provision of primary and secondary content.

These collections provide what I have termed "advanced access" to their materials. Here I demonstrate how Shelley-Godwin supports certain kinds of scholarly activity, through a complex of interrelated data models and tools. Throughout this explanation, I show how the attributes described in Table 9 – especially *items*, *interrelatedness*, and *diversity* – manifest the purposes of the collection. Palmer (2004) described how thematic research collections were evolving in their

support of scholarly activities, including deep and wide reading, searching, and other scholarly primitives such as annotating, comparing, referring, selecting, linking, and discovering (Unsworth, 2000).

The *conceptual* items at the heart of the Shelley-Godwin Archive are manuscripts of literary works. Primary sources are given primacy and distinction in this collection, as in all definitive-source collections; there are few secondary sources, in comparison to other kinds of collection. In the Shelley-Godwin Archive, secondary sources include some narrative introductions to and expositions of certain manuscripts and editorial choices. On landing pages for works, the collection points to related editions, chronologies, bibliographies, etc. Users can explore the collection as a definitive source for the manuscripts, and a reference source for related materials.

In underlying data models, however, the primary source items are represented as discrete page images and page-level TEI encodings. This granular representation enables a variety of manuscript representations that differ by their order and historical context: linear Work order, original Notebook order, etc. (See browsing options in Figure 10). The collection provides citation guides per page rather than per work, suggesting again the importance of the page-as-item.

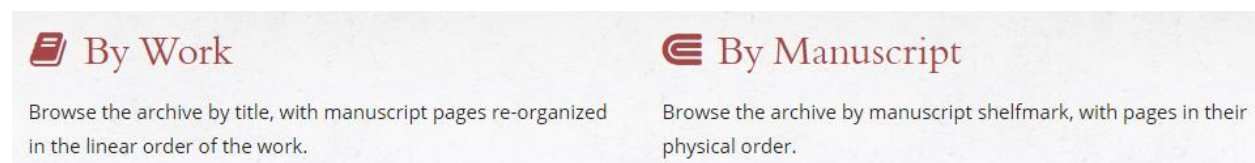


Figure 10. Different exploration options for manuscripts in the Shelley-Godwin Archive

Shelley-Godwin can be seen to support multiple kinds of scholarly activity, beyond searching and reading: including comparing different historical orderings of works, comparing page images against encodings, and identifying (and, concomitantly, citing and linking) works at the page level. Comparison among different versions of a text is also facilitated by two important aspects of the collection's TEI encoding scheme: the abilities to assert different authorial hands and to track revisions. (This is especially important in the Shelley-Godwin Archive, as the extent of Percy Shelley's contributions to Mary Shelley's work is a matter of scholarly debate.) Comparative views of different versions of items in the collection, enabled by the encodings, are presented by a homegrown Shared Canvas viewer. Figure 11 shows comparative views of the same snippet of text (clockwise from top left): the page image, the TEI-XML encoding, the fair copy

text, and the text with markup showing revisions in situ. Figure 12 shows an interactive viewer menu, enabling various modes of reading. Of course, primary sources in the Shelley-Godwin archive are interrelated primarily just by being collocated and jointly keyword-searchable. Figure 13, using the example of a keyword search for the theme “hero,” shows how this simple interrelatedness alone enables potential connection-making across works, and comparison across different versions of texts.

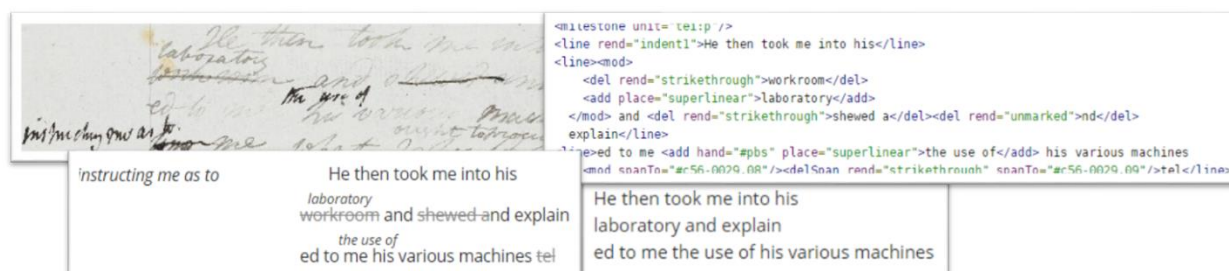


Figure 11. Comparative views of a single text in the Shelley-Godwin Archive

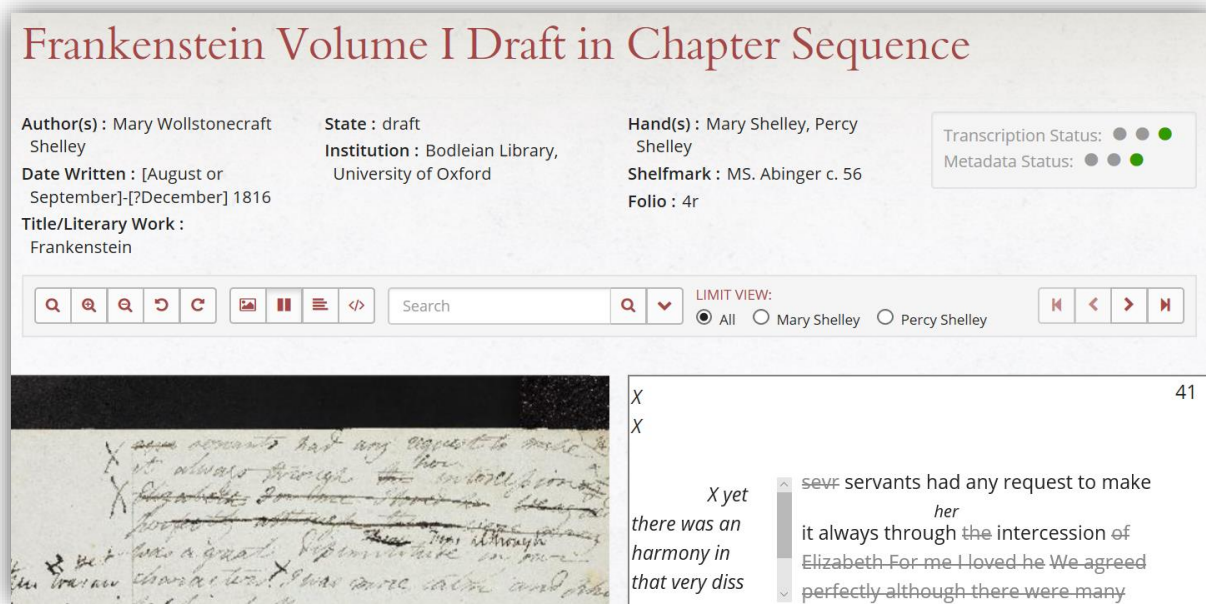


Figure 12. Multiple modes of reading in Shelley-Godwin Archive

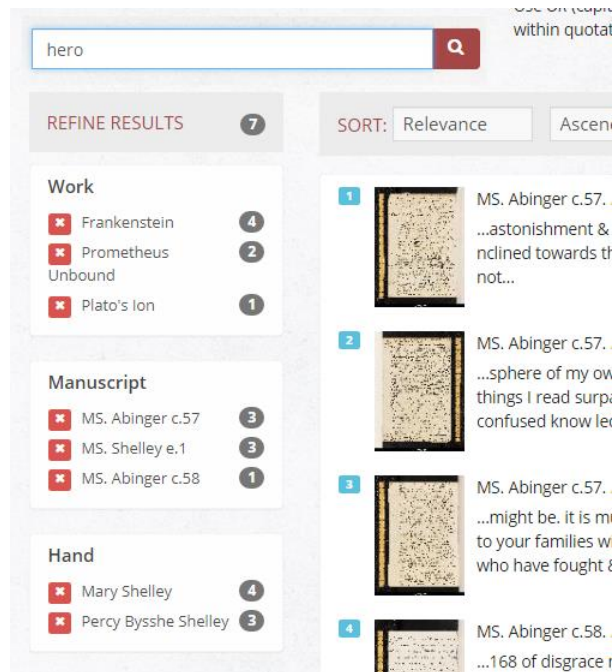


Figure 13. A search for “hero” across related works, manuscripts, authorial hands

The most important activity support of the Shelley-Godwin Archive is still aspirational, for now; the collection has built in facilities for a next generation of users to be able to actively forge relationships between works and other abstract entities:

the S-GA will thus function ultimately as a work-site for scholars, students, and the general public, whose contributions in the form of transcriptions, corrections, annotations, and TEI encoding will create a commons through which various discourse networks related to its texts intersect and interact.

However, the collection has laid foundations for definitive sources to be more readily manipulated even than in the current dynamic, but inflexible, comparative views. Manipulability – for annotation, for comparison, for collation – is pursuant to the aims of scholars who would employ those kinds of sources.

Definitive-source collections were the second most common type of collection in the sample, after Exemplar/context collections. We should not read too much into that, as the sample is still small, and heavily affected by the sources chosen for identifying collections.⁴⁷

⁴⁷ The commonness of the genre may be a reflection of many things: e.g., literary scholars’ interest in and comfort with the well-established and often overlapping genre of scholarly edition (as discussed in section 5.4., above), tempered by the expense of creation of a definitive resource; or the wave of related digitization efforts in the 1990s, as Nowvieskie recognized: “there were a lot of scholars who were familiar enough with principles of scholarly editing, unafraid of really big projects, that there was this late-90s / early-2000s blossoming of these things,” which

5.5.2. Exemplar/context collections

Like definitive-source collections, exemplar/context collections seek to enable common scholarly activities, including searching, reading, linking, and comparing, but these collections tend to place a stronger emphasis on discovery. These collections are built to help users discover unknown items and relationships among them. This is especially important when the theme of a collection is a subject around which extant special collections or archival sources do not exist. For this reason, exemplar/context collections often aggregate digitized primary sources from highly distributed and disparate source collections. For this reason, also, these collections are oriented toward a different notion of completeness, namely exemplarity.

The collections follow the metaphor of the “archive,” as in the case of many definitive-source collections, but they are often based on more nebulous themes. Themes tend to be concepts, events, places, or phenomena somewhere in between – all less well defined than, say, a given author’s oeuvre, or a single work. For example, “Uncle Tom’s Cabin and American Culture” revolves around a single work (*Uncle Tom’s Cabin*), but foregrounds exploration of that work’s influence on American culture. The latter half of that theme is difficult to circumscribe. Thus, the collection pursues a different kind of completeness than a collection built around providing definitive sources about or of the work itself. The collection mainly gathers and contextualizes *examples* of the work’s influence on culture.

Exemplar/content collections tend to be diverse. Sources of diverse kinds and on diverse topics (fiction, journalism, etc.) have potential to appeal to researchers from different disciplinary perspectives, seeking different kinds of evidence. In this way, these collections more closely resemble the interdisciplinary platforms described in Palmer (2004). In addition, these collections provide context to exemplars through the combination and interrelation of primary and secondary sources.

The Vault at Pfaff’s gathers a vast array of diverse items that are connected, sometimes indirectly and in different ways, to one historical-physical space – a bar called Pfaff’s – and by implication to an abstract network (or community), the New York bohemians. Through its interrelation of items, the collection almost asserts a kind of thesis: that there are important,

one participant termed, “early adopter projects and electronic archives..., especially in the 19th century British and American spheres” (P2). Of course, it could also be a reflection of the use of NINES as a source for the typological survey; a larger sample might change ratios of types.

perhaps undiscovered relationships between all of these thousands of authors and texts, revolving around the center point of Pfaff's, known to be a hub for the bohemian community in New York. Once again, the emphasis is on interrelating previously unrelated sources, and on providing a rich context for historical exploration, augmented by carefully curated bibliographic and biographic information.

Thoughts and Things No. IV

Clare, Ada. "Thoughts and Things No. IV." *New York Saturday Press*. 12 Nov. 1859: 2.
[Electronic Source Available](#)

Type: newspaper

Genre: commentary

Abstract:

In this column, Clare gives her thoughts on *Beulah*, which she criticizes as an "inane" see the *Marble Heart* and praises Laura Keene's acting style. She also discusses a p several members of the cast. Clare also shares her feelings on hot air balloons, recor previous edition of the *Saturday Press*, and criticizes "vain men" for "despising love".

People who Created this Work

[Clare, Ada](#) author

People Mentioned in this Work

[Gay, Getty](#) [pages:2]
 Clare recommends reading Getty Gay's "Waking from Illusions," a sketch found in the p

[Jefferson, Joseph](#) [pages:2]
 Clare calls Jefferson the "mainspring" of the most recent production of the play *Smike*. "regret[s] deeply that I ever to have seen him disgrace himself as Asa Trenchard" (2).

Figure 14. Contextual information and interrelation of sources in Vault at Pfaff's

Primary sources are usually the conceptual unit gathered by this kind of collection, and are foregrounded in search and browse. But more often than in definitive-source collections, exemplar/context collections mediate discovery and access with different kinds of interpretive affordance. For example, in the Vault at Pfaff's, the items, conceptually, are *works*. The great majority are primary sources, with a few original secondary sources (historical essays) in the mix. Primary and secondary sources are treated indistinguishably, except that secondary sources fall under different browse menus. In terms of their representation in the underlying database, the actual items here are bibliographic or personal records, which provide links to digital content, whether from external, third-party (aggregated) sources, or in the site's separate CONTENTdm server. Not all "items" (records) include links to digital primary sources; again, the priority here is discovery through and of the network. "People" entities appear to have item status in some senses: "people" are returned alongside primary- and secondary-source works in a keyword search, made

visible in the same way in the browse menu, and their landing pages are nearly identical, except that “People” are surrounded by more scholar-generated contextual metadata and narrative than are other items. For example, Figure 15 demonstrates how Vault at Pfaff’s foregrounds the navigation of different kinds of relationships between various entities, through hyperlinks embedded in item/person descriptions and a dedicated “Relationships” browser. Vault at Pfaff’s employs a controlled vocabulary for human relationship types, which have been manually asserted by editors of the project. Relationships between people and works are navigable as annotated links on relevant landing screens.

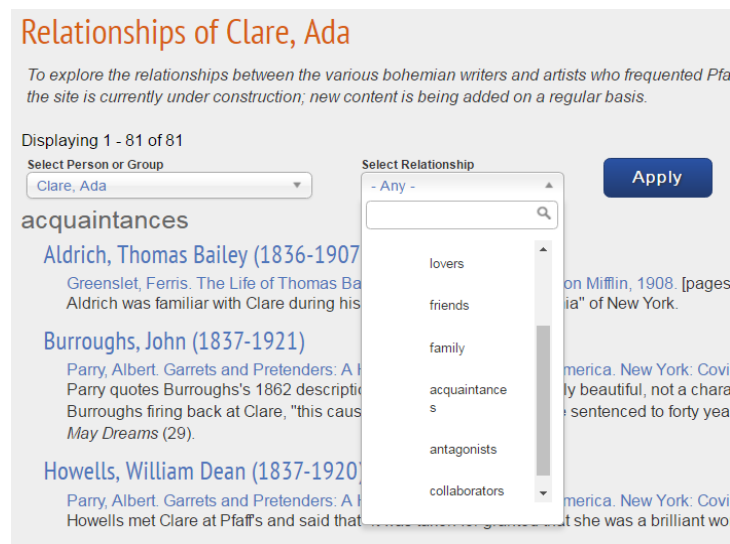


Figure 15. Relationship browser in Vault at Pfaff’s

While this was the most common type of collection, this type exhibited the most heterogeneity within itself. These collections all gather exemplars rather than definitive sources, but some collections are heavier on context than others. Many resemble conventional Omeka collections: a set of primary sources, gathered into a simple content management system, and augmented with relatively minimal metadata. Nonetheless, they have made an intellectual contribution just by the assertion of an item’s relevance to theme. In other cases, contextual information is a significant contribution, and the collection begins to resemble something more like a traditional scholarly product or secondary source. For example, Vault at Pfaff’s not only provides a couple long-form historical essays, one of which was published as a collaboration with

Lehigh University Press,⁴⁸ but also secondary materials in non-textual media: an interactive map of New York that highlights relevant places of business and residences, and an extended timeline with embedded images, both oriented toward teaching. Items in the collection are woven into biographical profiles, which include historical narrative based on evidence from the collection. Exemplar/platform collections offer varying layers of secondary scholarly contribution, beyond gathering items, on new themes, to facilitate research.

5.5.3. Evidential platform collections

Collections serve as evidential platforms, gathering primary sources and leveraging evidence – in the form of data – into new platforms for research and learning. These are, so far, the rarest kind of collection.

They tend to pursue more specific research objectives than the other two types. For example, Voting Viva Voce “explores the lives of the residents of two nineteenth century American cities: Alexandria, Virginia in 1860 and Newport, Kentucky in 1870. Alexandria was a commercial city based on slave labor; Newport was an industrial city based on immigrant labor.”⁴⁹ *Aquae Urbis Romae* is “an interactive cartographic history of the relationships between hydrological and hydraulic systems and their impact on the urban development of Rome, Italy,” beginning in 753 BC. These themes are more akin to the topics of historical monographs than to those of other thematic collections, but unlike monographs these collections generally refrain from making explicit, narrative arguments.

These collections gather primary sources but treat them differently, in part because they are driven by specific interpretive or analytic goals. The main goal is not to make primary sources accessible as such, but to identify, extract, and interpret data derived from sources. These collections are driven by the evidential potential of their sources, to be extracted and manipulated into serving as evidence toward specific research objectives, both by the creators of the collection and its users.

⁴⁸ <https://pfaffs.web.lehigh.edu/node/38098>

⁴⁹ <http://sociallogic.iath.virginia.edu/node/154>

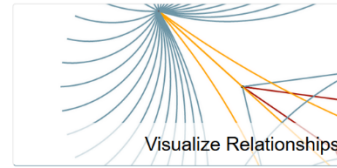
Bell, Harriet

Alternate Names: Beall

[Related Cases](#) [Related Documents](#) [Visualization](#) [Relationships](#)

Described in Case Documents As:

Sex: female
Age: 3 months September 14, 1835 [source]
Social Status: enslaved [source]



3 Related Cases:

- [Eleanora Bell v. Susan Armistead](#)
- [Harriet Bell v. Susan Armistead](#)
- [Mary Bell v. Susan Armistead](#)

Relationships:

- [Child Of Bell, Daniel](#)
- [Child Of Bell, Mary](#)

Figure 16. Landing page for Harriet Bell, demonstrating different abstract items and entities in network

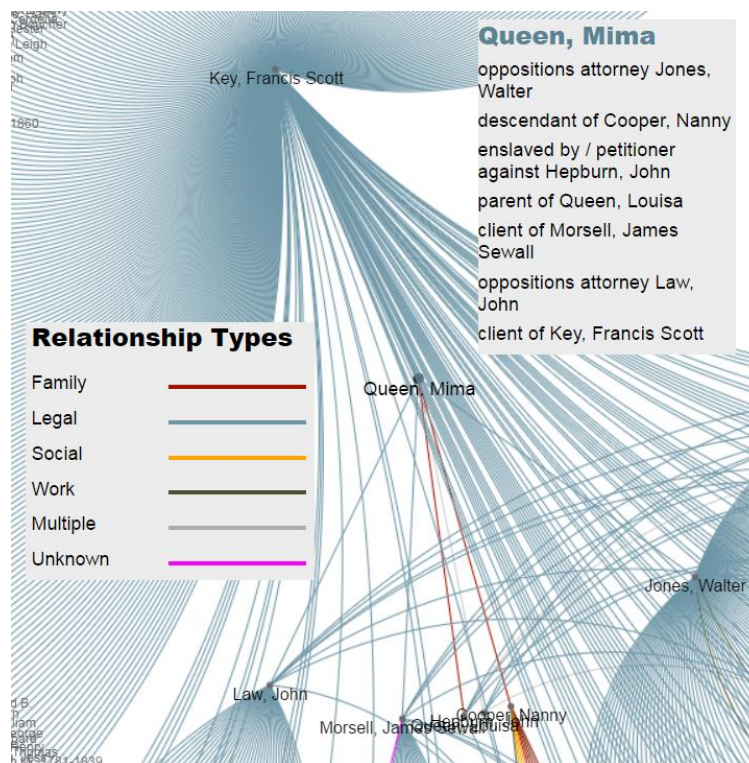


Figure 17. Visualized relationship network, O Say Can You See

The goal of O Say Can You See is to illuminate a multigenerational social network of African American families and related people in Washington, D.C., between 1800 and 1862. In so doing, the collection hopes “to make visible what has been invisible in the history of slavery, including the networks of relationships of the enslaved and free.”⁵⁰ To this end, the collection gathers criminal, civil, and chancery records, and freedom petitions from various court archives. These items, *documents*, are compound objects, represented simultaneously as TEI transcriptions

⁵⁰ <http://earlywashingtondc.org/about>

(with item-level metadata) and page images, all associated in the TEI file, and rendered as a single landing page with embedded images. However, these documents are not necessarily the central items of the collection. People entities, abstract entities derived (as cases are) from documents, constitute another set of items: browsable alongside cases, and searchable alongside documents. The representation underlying “people” items is a TEI personography. Subsets of people entities are interrelated in “family” entities, which are represented by annotated, genealogical trees (Figure 18).

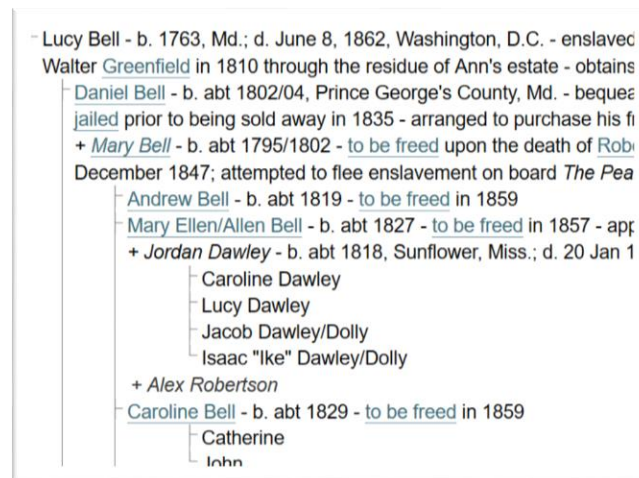


Figure 18. Bell family tree excerpt, O Say Can You See

Documents are interrelated into *cases*, another kind of abstract item. Cases are more immediately visible than documents, acting as a central browsing mechanism and serving to interrelate documents and people entities. Several cases receive significant scholarly annotation. Cases are not searchable directly by keyword, while documents (and named entities) are.

As in exemplar/context collections, interrelationships among items are central to the collection’s purposes. Unlike Vault at Pfaff’s, however, O Say makes relationships not through bibliographic (metadata) links, but through RDF encodings of data abstracted from the primary sources themselves. This collection interrelates its sources as a network of entities derived or abstracted from those sources: cases, people, stories, and families. Logically, a collection works by (1) offering a set of transcribed and encoded primary-source documents, which stand apart as historical record; and (2) from those documents extracting the data that serve to support the relationships among and analysis of the documents. These affordances serve to support further

analysis by enabling the discovery of previously invisible historical associations, and by interpreting and interrelating previously unrelated documents.

In O Say, these interrelationships are woven into a series of “Stories,” historical essays. These essays only represent a small fraction of the interpretive, historical narratives that this collection may serve to evidence. The narratives themselves are intended to be exemplary of the kinds of analysis the collection may support, like collection “highlights.”

These collections tend to be diverse, offering different kinds of data and sources that are potentially corroborative. Perhaps the earliest incarnation of an evidential platform collection is Valley of the Shadow. In an associated, pioneering multimedia publication, Thomas and Ayers (2003) describe the aims of the collection: “We investigate the problem of modernity and its relationship to slavery in these [two Civil War-era] communities by joining the tools of geography and cartography to those of social-science history, using detailed GIS to compare places and social, economic, political structures.” The collection gathers church records, letters and diaries, newspapers, statistics, maps, and census and tax records. Some are provided in document form; others are made available through database search. In this way, diverse kinds of historical record are brought into relation with one another, and brought to bear on a specific historical inquiry. Even as the collection serve their creators’ main argument, about the centrality of slavery to the Civil War and American modernity, it remains flexible to new kinds of use and interpretation along other lines of inquiry.

5.5.4. Making the case for these types

Types proved to be usefully discriminating in practice. I revisited the typological survey of collections, testing these new bins not only on the 98 collections already surveyed and categorized in the provisional typology, but on a new sample of 48 collections. The complete typological survey categorizes 145 collections (see Appendix A).

An ideal of typology is that types be mutually exclusive and internally homogeneous. However, the types offered here should be considered fluid. As Chapter 4 discussed at length, collections enact multiple purposes; so, collections may borrow characteristics or functions from across categories with different orientations. Nonetheless, every collection I surveyed (it can be argued) resembles one of these types. Appendix A provides the typology applied to a set of 145

collections. In the final iteration of applying and testing the typology on a fuller sample of collections, every collection fit into one of three categories, albeit some more readily than others.

Of course, there were significant variations within types, in terms of collections’ sizes, spatial and temporal coverage, forms, functions, etc. However, in terms of the attributes (purpose, theme, items, interrelatedness, diversity, and completeness) considered here, the homogeneity that does exist within each type is potentially useful. This is the best argument for the plausibility of these types.

By “useful” I mean that the attributes, which are categories of similarity or likeness between collections of each type, usefully (and, I hope, intuitively) manifest and explain each type’s overarching purpose. These attributes are indicators of how collections accomplish their scholarly goals – what they intend to contribute, and how they go about it.

A second argument for the plausibility of these types is their resonance with an existing, typological account of digital scholarship. Thomas (2016) articulated a typology of digital scholarship in the humanities, which distinguishes “thematic research collections” from “interactive scholarly works.” Table 10 is adapted from Thomas (2016).

Table 10. Excerpt of typology of digital humanities scholarship, adapted from Thomas (2016)

	Thematic research collections	Interactive scholarly works
Types of data	Heterogeneous, primary	Homogeneous, primary
Components	Schema, data models	APIs, scripting
Organization	Theme or subject	Hypothesis
Scope	Capacious	Tightly defined
Interpretive nature	Affordances	Query-based
Character	Open-ended	Procedural inquiry

Table 11 shows how Thomas’s typology can be extended with the one I have developed in this study. The typology I have described can be shown to reconcile with and substantially extend Thomas’s typology. I believe that his two categories of digital scholarship are two facets of a single, albeit eclectic genre (the thematic research collection). I align Thomas’s “interactive scholarly works” with my evidential platform collections, and his “thematic research collections” with my definitive-source collections. I also show how exemplar-context collections refine our sense of the diversity of contributions that collections make, taking on some of the characteristics that Thomas ascribed to collections generally, but which are not common to my definitive-source

collections, like heterogeneity and capaciousness of scope. Few changes are needed to basically represent two of my three types in ways that resonate with Thomas (2016). In Table 11, things highlighted in orange represent extensions to Thomas’s typology, drawn from my own. While our accounts differ on fine points – and differences may warrant further exploration in future work – the commonalities of our typologies suggest overlapping conceptions of the broad patterns of contribution that thematic research collections make to scholarship.

Table 11. Reconciling typology of thematic research collections with Thomas (2016)

	Definitive-source	Exemplar/context	Evidential platform
Types of data	Homogeneous, Heterogeneous primary	Heterogeneous, primary and secondary	Homogeneous, primary
Components	Schema, data models	Data models, scripting	APIs, scripting
Organization	Theme or subject	Theme or subject	Research objective Hypothesis
Theme Scope	Tightly defined: Capacious Author, work, archive	Less tightly defined: Event, concept, phenomenon	Tightly defined: Specific research objective
Interpretive nature	Affordances and encoding	Navigation and affordances	Query-based
Character	Open-ended	Open-ended	Procedural inquiry

The three types I have defined aim to offer sturdy handles for grasping an unwieldy genre, and for understanding what distinguishes thematic research collections from other genres of scholarly production. Types and their respective attributes are also indicators of, and offer a potential vocabulary for, the significant properties of collections that require perpetual maintenance to ensure their long-lived contributions to humanities scholarship. We will come back to this concept in the next chapter, on sustainability and preservation.

CHAPTER 6: SUSTAINABILITY AND PRESERVATION

This chapter concerns Research Question 2: What are the challenges, for libraries and related scholarly-publishing entities, in supporting thematic research collections as a scholarly genre? The chapter focuses on challenges confronting the sustainability and preservation of thematic research collections, strategies that have been employed or imagined for meeting those challenges, and current and potential institutional roles in sustainability and preservation efforts. Most of the issues discussed in this chapter – around challenges, strategies and roles – arose from the interviews, but the content analysis yields examples of certain challenges. The final section briefly considers the implications of the typology of collections for sustainability and preservation.

Sustainability and preservation are distinct ideas, but it is not clear how differently they are conceptualized or practiced in the context of maintaining thematic research collections. By *sustainability* I mean a collection's ability to remain fit for purpose over time, and to continue to grow and change. By *preservation* I mean, similarly, ensuring ongoing access to a collection over time through active management, but with the addendum that the collection may no longer be subject to growth and change, and that nonessential aspects of the collection may be allowed to fall away. Both terms are centrally concerned with maintaining the usability, understandability, and authenticity of digital objects over time (Smit, Van Der Hoeven, and Giaretta, 2011). Interview participants tended to conflate these terms. Given their professional roles and research priorities, most participants were interested in what Nowviskie termed "living" scholarly projects. In the case of projects that need to remain accessible, as one participant said, "It's not so much whether or not it's actively being enhanced or extended, as much as it is just what kind of technology is under there, and what has to be patched up when something shifts underneath" (P5). In addition, it seems that sustainability and preservation do not constitute a binary, in practice, but occupy different points on a spectrum.

In any case, the problem is urgent. It is clear from the interviews that the sustainability and preservation of digital collections over time is a significant source of anxiety for collection creators, and for people and institutions that have assumed responsibility for them.⁵¹ One

⁵¹ This sense of shared anxiety is confirmed by a recent, nationally scoped survey of 250 humanities scholars, conducted by the Publishing Without Walls project (PWW). The survey asked humanities scholars about their most valued forms of services and support for digital publishing. "Digital archiving and preservation" was deemed highly important and poorly supported in general, and scholarly requirements for these post-publication services is a subject of ongoing inquiry for that project. <http://publishingwithoutwalls.illinois.edu/>

participant observed the complexities of maintaining growing masses of idiosyncratic collections, each prone to unpredictable breakage:

...it's a pretty high burden, a pretty noticeable impact that these things make once they start to accumulate, because they aren't done in any standardized way, and they have all kinds of dependencies of a technical nature that are difficult to keep up with and maintain, and they rot in dangerous ways. So you can't really ignore them –P9

The contribution of this chapter is as complete a picture as possible (based on the interview data) of the challenges confronting the long-term management of thematic research collections. The conclusion of this investigation is not particularly optimistic. Attempts to systematize the treatment of collections (among other sorts of performative, digital scholarship) fail in library practice because collections are ever more complex – each collection demands unique treatment, indefinite curatorial attention, and careful assessment of and negotiation about commitments. As one participant noted:

We have no solutions for this. We have hardly even interim solutions. I'd feel bad about saying that if I didn't believe that everybody is in exactly the same or a worse situation that we are. Just all across the board, any direction you want to go with digital materials, nobody has any idea of what they're doing. –P5

When it comes to strategizing for the futures of these collections, according to another participant, “we're still in early days” (P8).

6.1. CHALLENGES

6.1.1. Interrelationships

The biggest challenge confronting sustainability and preservation is that thematic research collections are more than sets of discrete digital objects; they are dynamic, networked resources, in which the “connective tissue,” as Martin called it, may be as important as items themselves (if discrete items are even discernable as such). One participant noted a case in which the library assumed responsibility for a thematic research collection of Tibetan materials, but attempted to discard the scholar-generated, interpretation-rich indexing, in order to fit the primary sources into a preexisting content management system:

the way they looked to curate [the collection] was to grab all the images of the primary documents, and they were going to reformulate their own indexing into

it. And we said, well, no, this looks kind of like a normal index of Western documents, but it's highly interpretive ...and the library didn't ever get their heads around that. [E]mphasizing that the thematic repositories or collections have substantial metadata and interpretive materials interwoven is, I think, extremely important. –P5

Even in those cases in which items are the main contribution of the collection, and therefore the things that most need to be sustained (as in many definitive-source collections), maintenance will be complicated by how the items are represented through complex, interrelated data models and processes. Thibodeau (2002) described the difficulty of preserving compound digital objects and their interrelationships. This difficulty is witnessed even among the most apparently preservation-ready digital collections analyzed in this study. For example, the evidential platform collection *O Say Can You See* is primarily constituted of a set of ostensibly self-contained TEI-XML files representing archival documents – a relatively easy set of standard-conforming documents to maintain. However, a central contribution of the collection – namely, the scholar-generated relationships between people named in the legal documents – is offloaded into CSV and RDF files that interrelate entities abstracted from the TEI. Preserving the relationships between these objects is difficult. In a more extreme example, another participant described offloading editorial changes into processing, rather than changing data models directly:

[It's] possible to fix all kinds of problems in your data model by fixing them in the XSLT. I know this because I've done this many times. Because sometimes that's just the situation you find yourself in. In a perfect world, you would not do that; you would make sure that your texts and your data models took a very long view about future possible uses of your texts. The problem with putting too much in that, in the stylesheets, is that if those things ever get separated, you've lost a huge analytical contribution to those files. –P8

He noted the significant obstacle this poses for documenting data provenance and maintaining critical connections between collection components:

I'm not saying that everything can or should be in the TEI file. It's just that you need to be aware of the tight interconnectedness, the integration of purposes of these two things, the phenomena of the data model and the other, related phenomena of the stylesheet or the computational processes. –P8

Some collections take this challenge into account, and take pains to preserve some level of data independence. Planning for eventual migration into an institutional repository, Leon described

how Center for History and New Media collections tended to drive separation between the primary sources at the heart of their collections, and all the surrounding components: “Because we always imagine the stuff as migrating, [collection contents] would be easily separated from those interface issues.” Another participant observed: “If you want to preserve [the interface] as an interpretation, that’s a different sort of goal or task than keeping the data that’s inside the archive available” (P1).

Implicit in this challenge is another, common to digital preservation efforts: it is not always clear what aspects of a collection need to be sustained or preserved. Identifying the essential content, in resource-scarce situations, becomes a matter of conversation and negotiation between collection creators and would-be stewards. We will discuss this further under “Strategies,” below.

6.1.2. Short- vs. long-term priorities

Another sustainability challenge is that projects are propelled by short-term bursts of funding, and this shapes projects in ways that are incompatible with long-term maintenance. Of course, indefinite stewardship requires permanent resources – but even assuming those are available to a collection through a library, the pattern of short-term funding shapes projects in ways that complicate future maintenance.

Collecting projects may be compelled to grow rapidly, in pursuit of critical mass, as opposed to carefully, and with future maintenance in mind. For example, projects may forego documentation or the adoption of recommended but burdensome standards for encoding and description. Fraistat described a collective choice in the early days of Romantic Circles to prioritize “getting things up, getting things available” over adopting SGML, which had a “long production curve.”

Granting agencies usually prioritize research and development over other kinds of investment. Therefore, as we discussed under “Purposes,” collection creators develop collections in part as sandboxes for methodological research and experimental visualization and tool-building. Thematic research collections are not built purely in service to potential audiences; the process of their construction is research for their creators. Hybrid purposes, often born of funders’ priorities, affect every aspect of how collections are built, and therefore their ultimate sustainability. Participants reported an unflagging drive to innovate, to build upon, and to experiment with collections – an effect of research-oriented investment that leads to phases of distinct accretion of different data, data models, and affordances. Over time, a collection becomes “a gorgeous coral

reef of different approaches” (P7), which complicates efficient or systematic approaches to sustainability and preservation.

While sustainability is clearly a concern for collection developers, it takes a back seat to innovation and development. One participant said of preservation, “Yes, it’s a concern, but it’s not the driving concern...I didn’t want the concern about preservation to prohibit experimentation” (P3). Muñoz described a similar compromise made in selecting the Shared Canvas (now IIIF) data model for the Shelley-Godwin Archive, prioritizing an “innovative and interesting approach” over more familiar techniques despite the risk entailed in adopting cutting-edge approaches. Because the priority of collection creators is to conduct research, they eschew over-concern for the future: “I just think, I can’t worry about the unknowable future that much...What happens beyond me is not a problem I can solve or even deal with at all. I can do what I can do right now” (P3).

Experimenting often entails adding new kinds of data, new dependencies and interrelationships, and new kinds of affordances. One participant noted, “everything you add to your technology stack increases the complexity of supporting and sustaining it” (P1). But in fact, there is a potential conundrum here, which some participants recognized. As long as collections continue to innovate and find new ways to engage audiences, they are more likely to be sustained:

“I think the most important first point about Romantic Circles, when we started it, we very much felt like it was a research project, not a publication venue as such. ... We always wanted to find ways to keep on experimenting, keep on thinking down the road with it. And to the extent that a project remains alive that amount of time, you have to have that mindset.” (P2)

Collections tend to grow in spurts, distinct phases bounded by funding terms. Sometimes these phases are spaced apart by intervals of years. Between funding, collections may linger in stasis (though continuing to support research) or employ temporary, expedient measures to grow in the absence of dedicated support. For example, Muñoz described how the Shelley-Godwin Archive used classroom exercises and established a GitHub archive to continue work on collaborative encoding of the manuscripts as a “stop-gap measure” while the project pursued funding for a planned “participatory dimension” of crowd-sourced encoding and annotation.

Efforts to maintain collections are often inevitably reactive rather than anticipatory. Maintenance needs can arise suddenly when something breaks (sometimes catastrophically): “when it happens it’s usually some sort of emergency... the technology can shift out from underneath you and leave you with a totally unspecified amount of effort to try to move forward

or maintain something” (P5). In the absence of dedicated funding for the routine maintenance that collections require, some participants described foisting preservation work onto research grants as a sort of “rider”:

[Adding new materials] was the driver for getting the funding. And in the middle of all that we had to build a structure for those new materials. So reformulating the earlier stuff to go into it came – from the funder’s point of view – came for free, but at the same time we got essentially the funding to do the kind of migration that we needed to do but didn’t have the resources to otherwise –P5

6.1.3. Institutional contexts

Where a collection is born – for example, to what kind of institution or where within the hierarchy of a university system – may play a large role in determining its sustainability. Fraistat described how the sustainability of Romantic Circles at first benefitted from the lack of a University Press at Maryland, by positioning itself as an innovative publishing project, and positioning itself outside of any one department. Eventually, however, a shift in administrative philosophies left the collection at the mercy of department and college budgets, a shifting and uncertain space:

I’ve gone through now about four different provosts with different administrative philosophies, and we finally hit a provost whose philosophy was ‘devolve everything.’ So Romantic Circles devolved down to the College of Arts and Humanities level of support, which devolved it partly back into the English department, which meant its budgets were also potentially at hazard when the English department or the college budgets took big hits –P2

Of course, based on resources and staffing, different institutions have different capacities for maintaining and preserving collections. The DH centers represented in the interviews seemed to share a sense of responsibility for maintaining their own collections over time, rather than passing them off to a library, in part because collections represent the histories, contributions, and reputations of the centers:

both because MITH is a longstanding DH center and because it’s a space outside of the library’s collection workflow, MITH takes care of a lot of its own stuff for long periods of time –P6

Even within the university, which has an incentive to maintain its research investments, support may wane as the innovative potential of a project declines. Fraistat acknowledged that

Romantic Circles originally received university support because “the University at that moment wanted, through the humanities, to have some kind of technological breakthrough, something that they could point to as, ‘we’re ahead of the curve, and this is how we’re showing it.’” Considering shifts in support as institutional priorities change, Fraistat acknowledged that turning projects over from departments or centers to libraries could represent a viable solution to their sustainability: “maybe a particular university library would adopt projects like Romantic Circles, and keep them going, and help to staff them technologically – that’s one possibility I perceive.”

6.1.4. Collaborative workflows

There is a pressing concern among participants that the sustainability of collections depends heavily on the interest and engagement of individual creators. Interview participants suggested that collections are most sustainable when they are actively supporting ongoing research, and this is true not only for collections that have been centralized around one individual’s research interests. The need for a deeply engaged and invested leader extends to projects that represent long-term, collaborative efforts, including those situated within units or institutions designed to support and extend those collaborations. Fraistat observed this as a long-running concern:

I think the other big sustainability issue that hit every one of those ’90s projects, and I think is well exemplified with the Rossetti Archive, is how does a project survive its founder? And when is a large digital project done? ... [Sustainability] has a lot to do with the people who make something happen, and how long they themselves can stay in that game. Even if it’s not a question of they’re retiring, but it’s a question of them having other things they want to do in their careers.

Boggs imagined a solution in which, as a creator’s interest in the collection wanes, it could be handed off to new research teams with new, energizing research interests in the same body of evidence. At that point, the new team would assume responsibility for the collection’s maintenance. He acknowledged, however, significant social and technological obstacles to this strategy. I return to this under “Strategies,” below.

A related challenge is that it is not always clear when responsibility for the maintenance of a collection should shift from a creator to a stewarding institution, usually a library. The Roy Rosenzweig Center for History and New Media is one of the rare centers that has a working partnership with the library to preserve digital collections, though the commitment is mostly

around metadata and file preservation, rather than the preservation of more performative aspects of collections. Even in that case, one participant observed that it can be difficult to determine when it is time to shift the onus: “We have a standing agreement that they are the end point...But there’s never really any timetable on that” (P4). Another participant affirmed the difficulty of assessing project status, and suggested that the solution was close collaboration between collection creators and digital preservation experts:

[Preservation] remained a very closely connected department [to Scholars’ Lab], and interrelated, because ... it wasn’t always cut and dry whether [the projects] were done and ready for the library to migrate and preserve, and sort of embalm, or whether they were things that the scholar might still like to add to – they were just taking a hiatus from them and might want to come back –P7

In addition, participants suggested that libraries might be good for backing collection data up, but creators and centers might not be willing to relinquish control over the canonical version of the collection. For example, in grant proposals, another participant noted:

...we always included a letter of support from the library saying that they would be the eternal resting place of the digital collections... *in addition to* our own ongoing responsibility to keep the site live and available and backed up –P4

Sometimes sustainability is a matter of redressing collaborative workflows, rather than collections themselves. One participant noted that in a recent, large-scale project migration, the challenge was not only to migrate the collection itself, but to migrate active workflows for continued collection development:

Changing [workflow] *is* sustainability work. ... having a conversation about what the project wants, what the folks working on it like to do, want to do with it – that was sustainability work. And keeping their workflow intact in some ways, but just fixing some things that maybe weren’t working. –P1

6.2. STRATEGIES

Participants described several strategies for overcoming the challenges described in the last section. Most have described using numerous strategies, to varying degrees, and often in combination. The strategies are predominantly concerned with sustainability (as opposed to preservation), as that was the first interest of the participants in this study.

6.2.1. Levels of service

Preservation often starts with a conversation, especially when the collection creators and intended stewards are different parties. Negotiation about what commitments a library or other unit will make to a collection over time is an essential first step in preservation work. It is also a time-consuming and burdensome step. Some libraries have attempted to systematize these decisions by articulating levels of service and commitment. For example, the Sustaining Digital Scholarship (SDS) project offered various Levels of Collection to describe a library's commitment to preserving digital scholarship (I return to this below). The University of Victoria Libraries recently implemented a "Grant Services Menu," to help faculty understand the cost of library commitments to sustaining data and digital scholarship over time, and the "menu" is a centerpiece of the library's negotiations with faculty around stewardship of digital outcomes of their grant projects (Goddard and Walde, 2017).

These conversations, as SDS describes, can usefully revolve around the significant properties of collections. Significant properties are "those properties of digital objects that affect their quality, usability, rendering, and behavior" (Hedstrom and Lee, 2002); they function as guides to what aspects of digital objects must be preserved. One participant described how the concept of significant properties comes into play early in the collaborative process of collection-planning, but not specifically in regard to preservation:

...it comes up in passing, [when] we try to do a triage on how much effort something is going to take ... I'm not sure what blacksmiths call it when you're pounding the iron into steel – but anyway you are left with things that pretty much are important. We don't address it in very explicit ways, particularly with regard to curation, but it comes up indirectly –P5

Many times, however, collection stewards are dealing with an extensive backlog of collections that have been haphazardly maintained over time. In those cases, conversations and decisions about sustainability and preservation end up being largely retrospective, and especially burdensome. This is common in part because, as described under "Challenges," above, it is rarely clear when a collection project is done, so many linger in uncertain states, "cared for largely by benign neglect" (P6). When it comes time to make decisions about dispositioning digital collections, the initiative may come from different people in different roles – sometimes center staff, sometimes a faculty creator who is moving, or simply from someone who notices something is broken. Nowviskie described how Scholars' Lab sifted through a backlog of collections:

Our unit wound up doing a lot of stuff related to assessment of the collections, and in some cases we helped re-patriate collections and projects that had been built in the 90s, [where] the scholars had long since moved on [or] in some cases were graduate students at the time, so we in some cases decided that the project would be healthier migrating to a new university, where the primary person who cared about it was still active. So there were some that left and some that stayed, and some that got reworked and some that got fixed in amber, and it was all on a case-by-case basis. But it took many years...

One participant suggested that one strategy for sustainability would be to consciously and carefully lower the fidelity of projects for long-term maintenance. For example, he described converting dynamic sites into static sites through web scraping, once the need for dynamic interaction has passed. Implementing lower levels of commitment, in recognition of the significant properties of digital collections, is uncommon in practice due to social obstacles that Muñoz acknowledged:

...there's been a lot of effort on building digital humanities things in a way that's sustainable, as though we could design them from the beginning such that they will last longer. And to some extent, I don't think all of that effort is wasted, but I think it maybe led to too little thinking of letting people build what they will build and thinking about how to take versions of it, perhaps not at full fidelity, as the long-term representation of that work. And certainly there's been not a lot of explicit discussion about what levels of lower fidelity are useful and the dimensions to that. Mostly because no one wants to admit to the lowering of fidelity or the kind of non-perfection of a preservation process, for political reasons. –P6

6.2.2. Standards and documentation

Perhaps the most common strategy that collection creators described for sustainability was simply adherence to standards and best practices for data-model selection, metadata, and documentation. Shared standards and thorough documentation can preempt many sustainability challenges.

...there was a sense in which I was trying to show that I was doing a good job in my new job by thinking about the longevity of the project. And so, yeah, the TEI drove the decisions that way, even in some ways the IIIF [manifest], which is in itself a form of published metadata...This isn't just putting information into the application, which could go away. It's publishing that information as published metadata, which could have greater longevity –P6

A significant challenge to sustainability of complex collections, as described above, is the location of important information within processing scripts rather than within the data files themselves. In cases where processing is central to the understandability and usability of items, a collection may augment manual documentation by ensuring that automatic processes leave traces of provenance, automatically documenting the manipulations and changes to items as they make them. For example, Pytlik Zillig described a parity-checking apparatus he built to detect data loss during XSLT transformations of the 1,400 encoded text files of the Journals of Lewis and Clark collection.

It is incumbent on collection creators to document technical and design decisions, but there are few incentives for creators to do so thoroughly, as one participant observed: “I think we should feel an obligation to do that [documentation]. But there’s no real genre for humanities work or these thematic collections about documenting and sharing those sorts of things” (P1). Another noted that collection documentation is crucial to maintaining the authority and reliability of humanities evidence: “If we don’t clearly articulate the appraisal process or the selection process or the standards by which they’ve been assembled, people can’t use them, honestly. ... We should write much more clearly about it than we do” (P4).

6.2.3. Migration

As with digital preservation generally, migration – in several senses – is the main strategy for sustaining and preserving digital collections. Participants described different kinds of migration undertaken to maintain collection contents and infrastructures, including the migration of data into new data models (e.g., adopting standards as they become established), and moving data to new locations altogether. As mentioned above, one participant imagined the possibility of handing off collections to new research teams, mirroring patterns of open-source software development:

...instead of preserving and sustaining this for yourself for as long as you live, think about how you might document and structure the project if you were to say, I’m done with this, I want to actually hand it to somebody else who has energy and interest. –P1

He noted that collection-creators may be reluctant to relinquish credit for or control over their collections, which may represent years of scholarly endeavor. And, on the other hand, the prestige factor might limit uptake of extant collections:

There's not a whole lot of support, either financially or culturally, to keep working on somebody else's thing. Which I think is kind of disappointing, because it could work just fine. I'd like to figure out ways to do that for Neatline, because I want to anticipate a moment, sometime in the future, where we can't or don't want to work on Neatline anymore. But it's an interesting project and there are people that would like it –P1

Whether there are teams willing to take on and sustain whole, extant collections may depend in part on how capacious the collections are: how flexible those collections are for new kinds of use and new lines of inquiry. I come back to this theme in the next chapter, considering how flexibility, extensibility, and mobility factor into the capacity for collections to serve as platforms for research.

The prospect of a more diverse economy of moveable, flexible, repurposed and reusable collections – which also manage to accommodate established systems of scholarly credit and reputation-building – is an appealing one, and a subject of future work.

6.2.4. Community engagement

On a related theme, the effectiveness of a collection as a platform was seen by some participants as essential to its sustainability and preservation. The key to keeping collections around is to ensure their value to communities. This is more important to the persistence of collections than any technological decision, as one participant noted:

I know I've been quoted on some occasions saying, with regard to digital preservation, that the main thing is love, and love will find a way. It's a conviction that I have, that the projects that really resonate with their users, that have active communities that care about them, are the ones that are going to get migrated and preserved regardless of the challenges. So the bigger concern is not how do you structure these, how do you mark up your materials, how do you encode it all. It's really how do you create those kinds of community engagements that result in people squawking if the project goes away. –P7

Collection sustainability depends on engaging communities of interest, both user communities and development communities. Collections' concerted efforts to appeal to broad publics are essential: "initiatives like Shelley-Godwin and Romantic Circles...that cross the academy with larger publics are so important for the future of humanities... we've got to speak powerfully to the public about the importance of the humanities in ways that actually register." Perhaps even more essential is maintaining the interest and attention of local communities that can

respond to the maintenance needs of collections as they change, and as original sustainability plans and measures are inevitably compromised:

...the most fundamental part of any sustainability and preservation plan is making sure there are actually people around who want to continue doing that work [because] there's no plan that's going to survive ... when you actually have people and systems in place –P1

Part of engaging broad user bases might mean teaming up with other, thematically related collections to attain critical mass. Some participants suggested aggregation of collections as a partial strategy for increasing user communities, sharing infrastructure, and unifying preservation strategies for increased efficiency. However, aggregation may pose a procrustean solution: only limited aspects of collections can be aggregated – specifically kinds of data and metadata that have few interdependencies and pose the least challenges to sustainability and preservation.

While they do not solve the problem of functional preservation of complex digital resources, one participant saw the strongest bit-level preservation potential for the futures of thematic research collections in aggregate, networked repository systems for preserving the scholarly record, such as the Academic Preservation Trust⁵² and Digital Preservation Network.⁵³ These systems aim to respond to the weaknesses and variability of localized institutional repositories and data services, and the risks inherent in offloading data preservation to a commercial cloud. The Digital Preservation Network, for example, aims to developing a robust, large-scale, shared digital preservation infrastructure and set of services, supported by academic member institutions.

6.2.5. Redirection

A final suggestion for sustainability was to leverage the natural tendency of long-lived collections to change course over time. I discussed how collections can shift in purpose in response to factors ranging from the satisfaction of original goals and generation of new ideas, to shifts in funding and workflows. A couple participants suggested exploiting these changes to increase the sustainability of collections. For example, one participant suggested that as the research use of a collection declines in the eyes of its creator, its teaching potential may rise, and this is something that can justify maintenance of the collection:

⁵² <http://aptrust.org/>

⁵³ <https://dpn.org/>

...the research part maybe was biggest at the beginning, and is kind of tailing off, whereas the teaching part was smallest at the beginning, because there wasn't anything to show anybody, ... and it's growing as more students use it in class. Now that it's out there it can be incorporated into other pedagogical activities, so it's growing... in some ways I think of its pedagogical usefulness as core to the argument for its long term stewardship, more than the research –P6

Pivoting collections toward new purposes, such as from research-orientation toward teaching, resonates with the strategy of moving collections to new teams with new research interests. What would it mean to build collections in such a way that they may be pivoted to new lines of inquiry, or new functions, as intentional strategies for ensuring their sustainability? And how might this be reconciled with the notion that thematic research collections, by definition, are designed for deep inquiry on specific topics? These questions come up in Chapter 7, which explores how collections serve as platforms for research.

6.3. ROLES

At the individual and institutional levels, roles related to thematic research collections are complex and subject to negotiation. Roles within the system of scholarly communication at large become systematized and institutionalized only around established, well understood genres. When it comes to sustainability and preservation, how might roles be divided among the institutions that usually assume stewardship of collections – namely, libraries and digital humanities centers?

This study has focused on digital humanities centers as the main institutional wellsprings of thematic research collections, in part because the pilot survey of collections found little evidence that libraries have played a significant role in the creation of scholarly collections (except where a library contains a digital humanities center). Librarians are active in the digital humanities community, but this does not necessarily reflect established or sustained administrative or institutional support from the library; it may just reflect the initiative of individual librarians. One participant noted that librarians in service-oriented positions often enter into collaborations almost “in spite of or around the edges of their existing roles” (P6).

While library support for digital scholarship and digital publishing is on the rise (Posner, 2013; Bonn and Furlough, 2015), and thus we can expect more libraries to take more active roles in the collaborative creation of scholarly collections, participants were clear that there is a mission distinction between what research partners do and what service-providers do. When librarians are involved in collection-creation, it is usually in the capacity of partnering in the research and

development process. Muñoz suggested that maintaining the boundary between the library and the co-located digital humanities center (MITH) was important to the missions of both institutions:

What MITH chooses to do or creates is in some ways explicitly marked out as being outside the boundaries of the things the library is responsible for, in terms of its collections or operational activities. Part of the function that MITH serves is to be the *other*. –P6

While a center may be collocated with and have a strong relationship with a library, transferring onus for collections is still a matter of complex negotiation. Muñoz noted that sometimes MITH produced collections that look a lot like library collections, but that does not make them readily transferrable: “it would certainly be a negotiation to bring something across the boundaries of the library, even to prevent its demise” (P6).

All but two of the interview participants were affiliated with digital humanities centers situated at research universities. In each case, the center bears a different relationship to the university library. Table 12 gives a brief review of the administrative relationships (with number of participants representing each center in parentheses).

Table 12. Relationship of Center to Library at each interview participant's institution

CDRH (2)	Joint initiative of the University of Nebraska-Lincoln Libraries and the College of Arts & Sciences, located within library
IATH (1)	Independent research unit of the University of Virginia, located within library
MITH (2)	Jointly supported by University of Maryland College of Arts and Humanities and University of Maryland Libraries, located within library
RRCHNM (1)	Part of the Department of History and Art History at George Mason University, located in independent facility
Scholars' Lab (1)	Unit of the University of Virginia Library, located within library

Participants agreed that libraries have a significant role to play in the sustainment and preservation of digital collections. The library is seen as a locus for preservation potential. Most participants (6 of 9) reported having had at least one or two interactions with the library around the maintenance or preservation of thematic research collections; those were more or less successful. None reported having procedures or processes in place for systematic, ongoing sustainability and preservation. One participant confirmed this lacuna in libraries' service to digital scholarship:

I think it is generally true that libraries have not collected free-range, born-digital scholarship. I think they haven't done it because it's difficult and expensive, and potentially fraught in terms of the commitments the library is able to make and that the faculty member might expect. But it's part of the intellectual record of the university in the same way that you collect faculty papers or run an institutional repository. It seems you have a similar obligation with respect to these materials
–P9

Though libraries have in general struggled to systematize stewardship efforts, some have made fitful attempts to take up the charge. From 2000-2004, the University of Virginia conducted a Mellon-funded research project into the sustainability of digital humanities scholarship, called Sustaining Digital Scholarship (SDS).⁵⁴ That project aimed to offer policy recommendations for libraries, around collecting complex digital scholarly projects (Sustaining Digital Scholarship, 2004). We will come back to their recommendations below, under “Sustaining Digital Scholarship.” Despite the progress that project made on clarifying library commitments, Unsworth noted that implementation has lagged in the intervening years: “on my list of things to do now is go back and apply the policies that I developed years ago.”

There is a history of library commitment to the preservation of thematic research collections at RRCHNM. Leon noted that RRCHNM had a longstanding agreement with the library, that the library would be the “would be the eternal resting place of the digital collections,” and this was regularly written into data management plans submitted with grant proposals. The RRCHNM ethos, conducive to preservation, may be attributable to founder Rosenzweig's own scholarly interest in digital preservation (Rosenzweig, 2003; Cohen and Rosenzweig, 2005). However, the library's actual preservation commitments appear to be limited. The library commits to preserving item-level metadata and limited kinds of primary source items (discrete files) in its institutional repository, MARS.⁵⁵ As discussed above, this level of preservation – while critical for preserving unique primary sources – will not suffice for capturing what Martin called the “connective tissue” of thematic research collections, which may constitute their central contributions to the scholarly record. The only RRCHNM collection that is complete in MARS to date is the Papers of the War Department,⁵⁶ a definitive-source collection, the main contributions of which are (1) its unique content of 42,000 war department documents, digitized as page images;

⁵⁴ <http://dcs.library.virginia.edu/sustaining-digital-scholarship/>

⁵⁵ <http://mars.gmu.edu/>

⁵⁶ See the original collection at <http://wardepartmentpapers.org/>; and the repository collection at <http://mars.gmu.edu/handle/1920/8388>

and (2) the finely crafted metadata for each document. Institutional repository preservation at this level is suited to this kind of thematic collection.

When a digital humanities center is physically or administratively located within a library, there seems to be an almost unconscious reliance on the surrounding infrastructure to bear the weight of stewardship of collections. Jewell described how this sense of a surrounding, preservation-ready context relieved him to focus on the growth of the Cather Archive:

I don't have to constantly worry about [preservation] because there's an infrastructure around me that's thinking about this... I also have worked on the principle that we couldn't be guided only by what we knew could be preserved, that we had to create things of value to the community, we had to evolve those things that continued to have value.

One participant described that having partners in digital preservation at the library was essential, in part just to ensure that there is some person whose job it is to keep paying attention:

probably the most fundamental part of any sustainability and preservation plan is making sure there are actually people around who want to continue doing that work. Because ...there's no [sustainability] plan that's going to survive ...intact when you actually have people and systems in place. –P1

One significant advantage of engaging libraries as research partners in digital humanities projects is in moving the curatorial perspective upstream in the course of collection development, increasing the likelihood that collection creators will make sustainable decisions and document those decisions. One participant noted:

I think to the extent that we can become involved when data is being collected, if we can provide some needed something – whether it's data storage, or server space, or whatever – to the extent that we can find a way to be desirable partner early in the process (which means, less expensive than central computing, less prescriptive than people might expect us to be, etc.), then you at least have a foot in the door, you're aware of what people are doing on your campus, that opens the door for conversations about data journals, and documentation. –P9

Libraries can proactively seek engagement with projects to ensure their stewardship, even taking on the burden of hosting or other services as a way of getting a foot in the door, as one participant observed: “that's why we have these moldering piles of digital data on our servers, which -- that's the good news! They're not gone. They just need attention... But at least we have

them, because we provided server space for them” (P9). However, he was clear that the future of library engagement with collections will revolve around research more than service-provision:

I think one of the attractions of setting up digital humanities centers and creating a focal point for collaboration between librarians and faculty is that that can change the way faculty think about what librarians do and have to offer. ... I think that’s the future of the profession, really, is being a partner in the research enterprise –P9

The notion of fundamentally altering or broadening the sense of service provision was a common thread in the interviews. One participant suggested that by de-localizing their sense of service, libraries might stand a better chance of helping scholars to create more effective and engaging thematic collections. She argued that, while libraries “conceive themselves around service to their scholarly community, in terms of discrete faculty projects,” she encourages librarians to broaden their perspectives:

I’ve started encouraging folks to remove the word “faculty” from their thinking on these things, and start to position this service and collaborating to build projects like this as service to seekers and scholars more broadly, so that they get out of the mode that they’re building something that expresses one scholar’s point of view, and more into partnerships that allow them to create collections that many, many different scholars, interpreters, can operate off and can redesign. –P7

Participants also suggested that libraries, like digital humanities centers, have a role in educating faculty about the creation of sustainable collections, and in helping faculty acquire sustainable funding. The potential for outreach and training around creating sustainable collections is another potential benefit of libraries engaging digital humanities scholars as research partners, earlier in the process of collection development. One participant described taking on this role as he helps scholars conceptualize and design their digital projects: “I try to...teach how everything you add to your technology stack increases the complexity of supporting and sustaining it” (P1).

6.4. EXTENDING CURRENT FRAMEWORKS

In this section I return to two threads of prior and ongoing research on the sustainability of complex digital resources, and consider their applicability to thematic research collections. First, I return to the Sustaining Digital Scholarship project (SDS) to suggest possible directions for future work on extending SDS frameworks for significant properties and levels of collection. At the end

of this section, I examine the potential for other emergent models of alternative scholarly communication to suggest directions for the sustainability of thematic research collections.

6.4.1. Sustaining Digital Scholarship

The SDS project, as mentioned above, investigated strategies for sustaining digital humanities scholarly projects in the library. I focus here on SDS as one piece of the extensive literature on digital preservation and curation because SDS is the only project that has explicitly targeted the preservation of thematic research collections, including The Rossetti Archive, The Samantabhadra Collection, and The Salisbury Project, each of which was addressed in my typological analysis.

SDS undertook an extensive study of policy issues surrounding the sustainability of digital scholarship from a library's point of view, considering selection, collection, preservation, distribution, and deaccessioning. The main outcomes of their study comprise two frameworks, intended to help libraries and faculty negotiate about the library's level of commitment to the preservation of digital scholarship. For information on their methods and detail on their policies, see the SDS Final Report (Sustaining Digital Scholarship, 2004).

First, SDS offers a set of loose categories of significant properties of digital projects to help creators and library staff alike determine which pieces of a work need long-term preservation (Sustaining Digital Scholarship, 2004, pp. 9-10):

- Presentation (visual and design elements)
- Function (organization and control elements)
- Usage (properties related to intended use)
- Content (elements that hold or represent intellectual content)
- Relationships (intellectual and encoded relationships between elements)
- Navigation (structured or arranged paths)
- Development plans (long-term plans)
- Historical value (which elements will have historical value?)

I propose that the SDS significant properties framework can be refined and extended with a fuller list of attributes and definitions derived from the content analysis protocol developed and applied in this study (Appendix B). The content analysis protocol teases apart some of the SDS properties into more refined categories, reduces overlap among the categories, and offers more potential attributes and definitions grounded in the digital humanities literature that has grown and matured significantly in the years since the SDS report. While the lengthier protocol would be more burdensome to apply, it may also contribute greater clarity about what, precisely, needs to

be preserved or maintained in digital collections. Future work will aim to map the SDS significant properties to the content analysis protocol, and from that derive a refined set of significant properties of thematic research collections.

Second, the SDS report describes five levels of collection, rising from minimal to thorough preservation efforts, to help scholars and libraries negotiate the extent of their commitments (Sustaining Digital Scholarship, 2004, p. 11):

- Level 1: Collecting metadata only
- Level 2: Saving the project as a set of binary files and metadata only
- Level 3: The content can still be delivered as in the original (interrelationships intact)
- Level 4: Look and feel intact
- Level 5: The project is completely documented (as a complete artifact)

It may be possible, in future work, to relate different kinds of collections, or different kinds of collection *components*, to different requisite levels of preservation. For example, we can see a very crude alignment of the types identified in this study with potential complexity of preservation needs. Definitive-source collections revolve around providing definitive primary sources and metadata. If we can tease apart those sources – which constitute the central contributions of those collections – from the affordances that get built around them (e.g., comparative viewers or navigational functions that take advantage of advanced encoding), these collections may be amenable to lower-complexity levels of preservation. In contrast, evidential-platform collections, which base their contribution on processing and extraction from items, and the interrelationship of components – might usually pose a greater challenge to preservation.

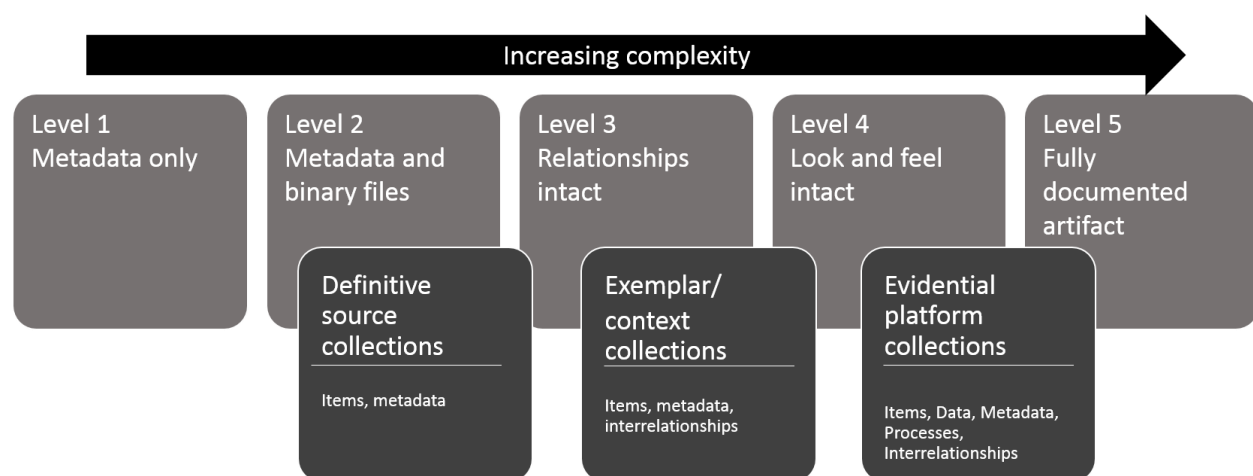


Figure 19. Rough illustration of how Types may inform preservation decisions

The rough association between the SDS levels of service and collection types made by Figure 19 is intended to illustrate how identifying types and their attributes can lead us toward more concrete conversations about the preservation needs of collections with different kinds of intended contributions to scholarship. Further work on this thread may help to improve what one participant (P5) decried as a lack of “well-formedness” around the objectives of digital preservation; the goals of preservation with respect to digital scholarship are rarely well defined. While SDS created a set of policies that have seen limited implementation, there is a need for ongoing research into how broader frameworks for digital preservation and data curation may accommodate or be adapted to thematic research collections as a distinctive genre.

Current strategies for collection sustainability – such as sustaining community engagement with ongoing experimentation and outreach efforts, and pivoting collections toward new uses to engage new communities over time – locate the responsibility for collection longevity with active collection developers and users. In cases where stewardship must transfer to the library, libraries may draw on extant frameworks for digital preservation and data curation. The central challenge will be ensuring their ongoing value as evidence for different kinds of inquiry, in accordance with the purposes of the collection. Evidential value depends on the integrity of the data and sources, which in turn depends on documentation of workflows and provenance. Existing research on documenting workflows and provenance has mainly considered scientific data (e.g., Stodden, 2010; Davidson and Freire, 2008). Documenting provenance for the use of humanities data and digital primary sources is an area of ongoing inquiry (e.g., Almas, et al., 2013).

How libraries decide to adopt frameworks of data curation will depend in part on how well they anticipate the future uses of digital collections and other forms of digital scholarship. Research into how scholars use humanities data collections and digital collections of primary sources is another area of ongoing work (Duranti, 2005; Palmer, 2005; Borgman, 2012; Padilla, 2016).

Scholars and libraries may seek to ensure the long-lived contributions of digital collections by finding opportunities to connect their research outcomes to larger data networks, especially through linked data and networked infrastructures for collaboration (Edmond, 2016). Humanities data sharing, particularly through linked data ecosystems, was a potential sustainability strategy raised in the interviews. How linked data and data sharing generally can contribute to the sustainability of humanities research is an area of active inquiry (Hoekstra, et al., 2016).

6.4.2. Research Object and Enhanced Publication data models and management systems

Some of the most significant movement toward creating systems for managing and communicating research data and compound digital objects stems from research and development of enhanced publications and research objects (see section 2.3.3 for definitions and contextual information about these genres of production). In this section I describe parallels and differences between research object models, enhanced publication data models, and the structures of thematic research collections. I then consider repositories and management systems built for research objects and enhanced publications, and their potential suitability for the long-term management of thematic research collections.

Both enhanced publications and research objects have been subject to modeling efforts, which have achieved some significant standardization of their representation. Of course, no common data model exists (in practice) for representing thematic research collections as aggregates, in part due to the challenges described in 6.1, and in part because thematic collections hardly constitute a coherent genre with unified structural features. Enhanced publications and research objects, however, are commonly modeled as compound digital objects. Seminal examples of enhanced publications, much like some thematic research collections, relied on hyperlinking and file naming schemes to represent and manifest relationships among their components (Holl, 2012). At this point, however, the most common data models for both genres rely on the concept of aggregations as formalized in the Open Archives Initiative Object Reuse and Exchange standard (OAI-ORE)⁵⁷ (Verhaar, 2008; Van de Sompel et al., 2009; Bardi and Manghi, 2015). Beyond being modeled using extensions of OAI-ORE, these resources are often also made available as linked data, and rely on persistent URIs and DOIs to facilitate citation of objects and their components (De Roure, 2014). The Research Object Ontology,⁵⁸ for example, is an OAI-ORE extension; and Figure 20 gives an example of an enhanced publication data model relying on OAI-ORE aggregates (Bardi and Manghi, 2015). Enhanced publications have also been realized using various proprietary data models (for use in commercial publishing systems), and EPUB3 has been shown to be a strong candidate for meeting the functional requirements of enhanced publications (Heyvaert et al., 2015; Sigarchian et al., 2014).

⁵⁷ <https://www.openarchives.org/ore/>

⁵⁸ <http://www.researchobject.org/specifications/>

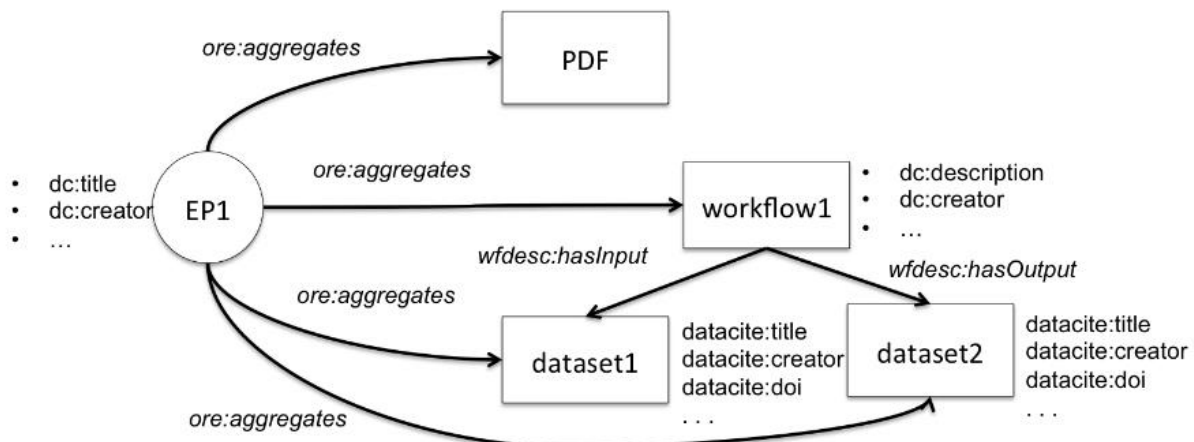


Figure 20. Example of an Enhanced Publication model, from Bardi and Manghi (2015)

The OAI-ORE is likely to be adequate as a basic model for representation of thematic research collections. In an analysis of enhanced publication data models, Bardi and Manghi (2014) identified five recurring classes of parts:

- Embedded parts (e.g., supplementary material files)
- Structured-text parts (hierarchical text with identifiable sub-components)
- Reference parts (e.g., URIs to external objects)
- Executable parts (e.g., software and data)
- Generated parts (dynamically generated components that may change based on inputs)

Conceptually, these classes readily contain the items and – perhaps to a lesser extent – the more complex “connective tissue” identified in my content analysis of thematic research collections. In practice, a model built around these classes may struggle to accommodate the full functionality of the interactive aspects of collections. Search functions, interpretive browsing functions, comparative viewers, dynamic social networks, interactive models – often the pieces of collections that determine the possibilities of their use for research – would likely challenge the “executable” and “generated” classes of components as they are defined for more limited application in enhanced publications. In addition, Bardi and Manghi found no support for these parts of enhanced publications in extant management systems. Nonetheless, this account, which is ultimately implemented using OAI-ORE, resonates with what is distinctive about thematic research collections and why they do not “fit” well into common repositories for managing other, more

conventional scholarly products. While thematic research collections have not yet been subject to OAI-ORE modeling as aggregates (to my knowledge), there have been applications of the model to traditional and digital archives (Ferro and Silvello, 2012; Guéret, 2013) and to *worksets*, which are like thematic research collections but oriented toward computational research (Green, et al., 2014; Jett, 2015).

Bardi and Manghi (2015) provide a list of requirements for management systems for enhanced publications, which may serve well as a starting point for managing thematic research collections in the aggregate. A repository or management system capable of accommodating complex thematic collections, without undermining significant differences across collections, could help systematize their institutional management and make long-term sustainability efforts more feasible.⁵⁹ According to Bardi and Manghi, an enhanced publication management system must:

- Support the integration of heterogeneous content from dynamic data sources
- Support the adoption of different storage back-ends
- Enable sharing via standard protocols
- Support portability of publications and components
- Support the enrichment and curation of publications
- Enable the definition of customized enhanced publication data models
- Offer languages for enhanced publication definition, manipulation, and access
- Support the addition of new domain-specific functionalities

The requirements for integration of heterogeneous content from distributed sources, and for pervasive data-model agnosticism and customizability, are particularly important when we consider their application to management systems for thematic collections. Interview participants, in describing previous efforts to aggregate diverse collections, testified to the necessity of balancing efforts at standardizing across collections and digital humanities projects against accommodating creativity and difference in this highly diverse genre of production. (In addition, the requirement to support “portability” foreshadows the final chapter of this dissertation, which reflects on characteristics of thematic research collections as flexible, extensible, and mobile

⁵⁹ Research object management, on the other hand, is increasingly oriented toward workflow; thus, repositories are often integrated with research infrastructures (e.g., Assante, 2015). It is hard to imagine this kind of repository model being useful for thematic research collections as they stand, given the current lack of common research infrastructure.

platforms for research.) As Van de Sompel et al. (2009) note, while OAI-ORE has demonstrated its usefulness as a basic model for representing enhanced publications, implementation relies on community-defined vocabularies for expressing precise relationships among resources in different domains. The same will be true for thematic research collections, which come from every humanities discipline. Chapter 7 describes future work on this front, which will address the application of OAI-ORE to thematic research collections, and the adequacy of existing ontologies, including those arising from enhanced publications and research objects, for extending the OAI-ORE to the representation of collections of this sort.

CHAPTER 7: COLLECTIONS AS PLATFORMS⁶⁰

In this concluding chapter I summarize and relate the outcomes of this study to my research questions:

(R1) What are the defining features of thematic research collections as a scholarly genre?

(R2) What are the challenges, for libraries and related scholarly-publishing entities, in supporting thematic research collections as a scholarly genre?

I proceed through these questions in reverse order, beginning with the practical challenges and opportunities that confront the creators and stewards of thematic research collections, and considering in particular the role of the library. Finally, I return to the defining features of thematic research collections, to begin to build a basis for future work on understanding the evolving shapes of digital scholarship and the contributions they make to humanities research.

7.1. CHALLENGES AND OPPORTUNITIES FOR LIBRARIES

This study suggests that the biggest challenges of thematic research collections for libraries lie in ensuring their sustainability and, ultimately, their preservation. Digital scholarship, in general, confronts the same sustainability challenges. This study identified four factors that compromise the long-term contributions of thematic research collections:

- **Complex interrelationships:** Collections are more than sums of their parts. The “connective tissue” that interrelates items within a collection, and which relates items to contextual information, represents a central challenge to sustainability. The parts of collections that serve to interrelate items are not often self-contained, and tend to be based on processes rather than static content, and yet they are often where the central contributions of a collection lie.
- **Short vs. long-term development:** The tension between short- and long-term development priorities may lead to compromises between immediate development needs and the long-term health of the collection. Preservation was not reported as a common concern for participants in this study, whose immediate interests in thematic research collections are research-oriented, and center on living scholarly projects. Phased growth and the

⁶⁰ Pieces of this chapter were first published in Fenlon (2017).

continuous pursuit of innovation may lead to the accretion of multiple data models and processes, with complex interrelationships that are difficult to maintain. Creators may favor rapid growth over the implementation of burdensome, sustainability-oriented standards.

- Institutional contexts: The sustainability of collections depends heavily on their institutional contexts, especially where they are born within the hierarchies of university systems. The nature of funding and support that spawn collections plays a determining role in how long collections may last. While research support can be short-lived as it focuses on short-term innovation, research-oriented institutions – like digital humanities centers – often find themselves performing stewardship roles.
- Creator dependence: Collections are highly creator-dependent, despite that they also usually represent large-scale collaborative efforts. As a creator's interest in or use for a collection wanes, there is a need for strategies for handing off the responsibility for a collection, either to a memory institution or to new research collaborators. In addition, sustainability strategies must consider not only collections themselves, but often the ongoing and active workflows still contributing to their growth.

For the most part, libraries are not involved with the co-creation of thematic collections, which tend to originate with independent scholarly efforts or in digital humanities centers. Where libraries are involved, they confront the same challenges that participants reported in digital humanities centers: conceptualizing a collection and its goals, obtaining resources and building collaborations, negotiating roles, compromising between different stakeholder expectations and the urgency of innovation.

The rest of this section describes the implications of this study for two communities with a stake in the future of thematic research collections:

- Academic and research libraries, in supporting the development, sustainability, and preservation of thematic research collections.
- Standards-making communities, including the Dublin Core community, in understanding the limits of extant collection-level descriptive schemas for representing thematic research collections.

7.1.1. Implications for academic and research libraries

This study identified several strategies for confronting the sustainability and preservation challenges presented by thematic research collections, including:

- Negotiation between stewards and creators over the significant properties of collections and the levels of commitment each is willing to make to the life of collections;
- Anticipating sustainability challenges by employing established standards and thorough documentation;
- Migrating collections as needed;
- Prioritizing the engagement of multiple audiences or communities to ensure interest and resources over time; and
- Pivoting collections toward new research objectives, or toward teaching and other purposes, to sustain community interest and value.

These are not novel preservation strategies for complex, digital objects. The last two, however, stand out as social strategies for improving the odds of collection sustainability. For collection developers, social strategies for sustainability are more important than technical strategies for the preservation of collections because developers intend collections to be active, growing hubs for collaboration and research, for as long as possible. Thematic research collections are set apart from many digital scholarly projects in the humanities by being platforms for new research, rather than serving mainly to express or disseminate already completed research. The interactive aspects of collections may complicate their preservation; but they may benefit collections in terms of sustainability by opening the possibility of engaging and renewing active use and development communities in the growth and maintenance of the collection over time.

Libraries have made forays into maintenance and preservation of thematic research collections, but systematic strategies remain elusive. The Sustaining Digital Scholarship project identified a set of significant properties common to thematic research collections and other digital humanities projects. The collection attributes identified for the content analysis protocol employed in this study correspond to and extend the properties that the SDS project identified. While the acquisition of thematic research collections into library collections is not yet a common practice, this study broached this prospect from two perspectives, stewardship and access:

- *Stewardship*: A library may physically collect a thematic research collection, and in so doing assume responsibility for its indefinite preservation (until deaccessioning, if that ever happens). However, most libraries are ill prepared to collect thematic research collections at full fidelity, maintaining the interfaces, affordances, and interrelationships that these scholarly resources contribute. Most aim to integrate collections into existing preservation streams, through deposit into institutional repositories.
- *Access*: Libraries may virtually collect thematic research collections as open-access resources by adding them to their catalogs and other discovery platforms. The goal of this approach to collection would not be to support sustainability or preservation (directly), but rather to enable broader discovery and use. This practice is largely uncommon, except for the most well-known and established collections. However, libraries increasingly virtually collect open access resources of other kinds, so this may just be a matter of collection creators taking the initiative and performing the necessary outreach. The efficacy of this kind of collecting will depend on rich collection description.

The sustainability and preservation strategies identified in this study relate directly to the roles of libraries discussed in chapter 6.3., as outlined in Table 13.

Table 13. Strategies for sustainability and preservation related to library roles

Strategies	Corresponding roles for libraries
Standards and documentation	Proactively engaging in research partnership to move involvement upstream
Determining variable levels of service	Negotiating levels of library commitment
Migration	Determining and negotiating transfers of responsibility
Community engagement	Broadening sense of service-orientation
Redirection	Proactively engaging in research partnership to move involvement upstream

Librarianship increasingly entails not only serving the overarching academic mission of an institution, but also actively partnering with faculty to produce digital scholarship. Participants in

this study suggested that partnership, as opposed to service, will characterize the future of librarianship in support of digital scholarship. This shift may be more transcendent, even, than a shift in patterns of service or collaboration: it may entail a shift in how libraries perceive their service communities. One participant suggested that engaging in research partnerships makes it necessary for librarians to broaden their service orientation, from supporting the academic mission of their university to serving broader communities, including the academy as a whole and the public. Libraries have already commenced research partnerships across disciplines by creating and staffing embedded interdisciplinary research centers, or adding staff with information expertise -- such as data curators and project managers -- in relevant roles on faculty research projects (Palmer and Fenlon, 2017). Assuming the stance of partnership (over service) may have the added benefit of moving sustainability and preservation upstream in the development process of digital projects, as well as affording projects an institutional context that supports long-term stewardship.

As libraries ramp up efforts to collaborate on digital scholarship, many of the challenges that beset thematic research collections may extend to the stewardship of other genres of scholarship in the humanities and beyond. Progress on generalizable solutions to these challenges depends on a foundational understanding of thematic research collections among other emergent genres of scholarship. This study contributes to that end by identifying the defining features of thematic research collections: what collections aim to contribute to scholarship, and how they go about it.

7.1.2. Implications for collection description

Sustaining or preserving the contributions of collections, particularly as platforms for new research, rely on effective documentation and description of collections and their provenance. While this study detailed the ways in which creators use or derive benefit from their own collections, especially in the process of collection development, the use and reuse of collections by other scholars and communities is not well understood. My future work will assess scholarly use and reuse of collections, especially seeking evidence of the use of collections as platforms for building new, born-digital interpretive scholarship.

Effective discovery and use of collections as sources of data and evidence for new scholarship relies on their thorough documentation, particularly documentation of data models, data provenance, and the decisions that have determined both. It is not clear whether existing descriptive schemas for collections and related digital objects (such as data sets) will be adequate

for representing thematic research collections. My next phase of work will assess the Dublin Core Collection Applications Profile (DC-CAP), the most common schema for library collection-level description, to discern its relevance to thematic research collections. The DC-CAP aims to support the discovery, identification, and selection of collections. However, its functional requirements may not align with scholars' intended uses of thematic research collections. For example, scholars seeking to engage in close examination or reading of sources will rely on significantly different kinds of description than scholars seeking to explore networks of connections between items, or seeking to undertake fundamentally repurposive computational analysis of underlying data. Thematic research collection description must support the evaluation of the evidential and analytic potential (Palmer et al., 2011) of a collection's items for research; in addition, it may also support intensive review and scholarly evaluation, for credit and reputation-building.

The differences between the collection analysis protocol employed in this study and the DC-CAP suggest significant differences between the aspects of collections that interest scholars and those that are readily represented by existing collection-description schemas. While the DC-CAP figured into the protocol development, attributes drawn from the literature on humanities scholarship and evaluation substantially extended the core properties of the DC-CAP, with detailed attributes specific to the context of digital humanities scholarship. Table 14 gives a preliminary assessment of the sufficiency of DC-CAP elements to represent the attributes of the content analysis protocol. Many of the attributes identified in the content analysis have no ready correlation in the DC-CAP (except perhaps in the Description element, which may serve to accommodate miscellany); those with some correspondence are still not sufficiently represented by DC-CAP elements for the most part. My future work will extend this analysis, and consider how collection-description practices – in the context of discovery, identification, selection, and use – may benefit from the content analysis protocol used in this study.

Table 14. Preliminary assessment of mapping content analysis protocol attributes to DC-CAP elements

Content analysis attribute	Potential mapping to DC-CAP	Qualification
Theme	Subject [dc:subject] ; Spatial Coverage [dcterms:spatial] ; Temporal Coverage [dcterms:temporal]	DC-CAP elements are not sufficient to describe Theme
Purposes	N/A	
Impact	N/A	
Creators	Collector [dc:creator]	
Audience	Audience [dcterms:audience]	
Documentation	N/A	
Provenance	Custodial History [dcterms:provenance] ; Accrual Method [dcterms:accrualMethod]	DC-CAP elements are not sufficient to describe Provenance
Related collections	Associated Collection [cld:associatedCollection]	
Related projects and publications	Associated Publication [dcterms:isReferencedBy]	DC-CAP element is not sufficient to represent Related projects
Review	N/A	
Funding	N/A	
Developmental stage	Accrual Periodicity [dcterms:accrualPeriodicity]	DC-CAP element related to but not same as Developmental stage
Host or publisher	Is Located At [cld:isLocatedAt]	DC-CAP element related to but not same as Host or publisher
Rights	Rights [dc:rights] ; Access Rights [dcterms:accessRights]	
Sustainability and preservation plans	N/A	
Method	Accrual Method [dcterms:accrualMethod] ; Custodial History [dcterms:provenance]	DC-CAP elements are not sufficient to describe Method
Items	Item Type [cld:itemType] Item Format [cld:itemFormat]	DC-CAP elements are not sufficient to describe Items
Diversity	N/A	
Size	Size [dcterms:extent]	
Narrativity	N/A	
Quality	N/A	
Language	Language [dc:language]	
Completeness	N/A	
Density	N/A	
Spatial coverage	Spatial Coverage [dcterms:spatial]	
Temporal coverage	Temporal Coverage [dcterms:temporal]	
Interrelatedness	N/A	
Data models	N/A	
Navigation	N/A	
Infrastructural components	N/A	
Interface design	N/A	
Interactivity	N/A	
Interoperability	N/A	

Table 14. (cont'd.)

Openness of components	N/A	
Identification and citation	Collection Identifier [dc:identifier]	DC-CAP element is not sufficient to represent Citation aspect of this attribute
Modes of access and acquisition	Is Accessed Via [cld:isAccessedVia]	DC-CAP element is not sufficient to represent Modes of access and acquisition
Accessibility	N/A	
Flexibility	N/A	

7.2. DEFINING FEATURES OF COLLECTIONS

One goal of this study was to enhance our sense of the intended epistemic outcomes of thematic research collections. Typological analysis and the development and application of the collection analysis protocol yielded a set of defining features of thematic research collections, building on the basic properties entailed in our definition: a collection of primary sources gathered by scholarly effort to support research on a theme. The collection analysis identified 38 properties that thematic research collections variously manifest or implement. Through typological analysis, several of those properties were shown to weave into common patterns, each oriented toward making a distinctive kind of contribution to scholarship.

By narrowing in on purpose as a central, defining feature of thematic research collections, I identified three types of collection, defined by how they manifest a certain constellation of properties: purpose, theme, nature of items they collect, how they define completeness, the interrelatedness of their items, and their diversity. These properties seem to follow from the basic purposes of the collections. The three kinds of collection identified by my analysis are basically differentiated by their central or defining purposes, which are manifested in the different ideals of *completeness* toward which these collections are developed:

- 1) Definitive-source collections are designed to collocate authentic, high-quality, and value-added sources, and they strive toward a kind of completeness that I have described as *definitive*: they are exhaustive with respect to their themes.
- 2) Exemplar/context collections aim to interrelate and contextualize diverse sources, and revolve around the relationships between items. They strive toward a kind of completeness I have characterized as *exemplarity*: they gather representative examples of sources that illuminate a theme, but make no claims to definitiveness.

3) Evidential platform collections aim to aggregate sources and then remodel them or extract evidence from them, to enable new kinds of interpretation and analysis. These collections strive toward *evidential sufficiency*: gathering enough evidence to support their specific research objective or answer a specific research question.

It may be that it is most useful to think of the “types” I have described as articulating three common patterns of contribution to scholarship, rather than truly, ontologically distinct kinds. We might think of these as constellations of attributes that seem to follow from one another, in pursuit of different kinds of contribution and completeness. Collections with multiple purposes might draw from across these types to accomplish their unique goals.

These types have potential practical implications for the evaluation, description, and preservation of thematic research collections. Guidelines for the evaluation of digital scholarship tend to compensate for the heterogeneity of digital scholarship by asserting that each project must be evaluated on its own terms, and in accordance with its own objectives. At the same time, there are calls for genrefication in digital humanities, in part to improve the process of scholarly evaluation and comparison (e.g., Thomas, 2016). These types give us a sense of the breadth of purposes that can motivate collection development, and provide a more stable scaffolding of language for use in describing and evaluating the contributions of collections, for purposes such as scholarly promotion, review and recommendation, and preservation assessment. They could be used, for example, to clarify the language of library recommendations or commitments to sustaining and preserving digital scholarship. In particular, the types point to significant properties of collections that are first priorities for stewardship efforts:

- In definitive-source collections: Primary sources must be sustained or preserved with sufficient contextual and provenance information to ensure authenticity.
- In exemplar-context collections: If primary sources are maintained elsewhere, it may be sufficient to sustain or preserve rich metadata for those sources, including permanent identifiers. It may be necessary, however, to preserve the relationships among items and supplementary materials and functionalities. These relationships, in implementation, may take the form of RDF documents, processing routines, schemas, or other instantiations of different data models.
- In evidential platform collections: Depending on the research objective of the collection and the methods employed in manipulating sources to answer it, sustaining or preserving

this collection may mean prioritizing curation of data derivatives of primary sources, rather than sources themselves, including workflow and provenance documentation.

The next phase of research will include a close, conceptual analysis of the notion of *completeness* with respect to thematic research collections, and its relationship to other properties of collections, such as *theme* and *interrelatedness*. In addition, the next phase of analysis will empirically test this concept by assessing whether the three kinds of completeness identified by this study can be considered representative or comprehensive of the ideals of completeness for digital collections in the humanities and in other contexts. The goals of the next phase of research will be to (1) articulate how these definitive properties contribute to making scholarly collections into uniquely supportive contexts for research; and (2) to examine whether these properties could be usefully applied to collection development more generally. This study has drawn from prior and ongoing work on the creation of *worksets* in computational environments (Fenlon, et al., 2014), and the analysis and identification of topical pools within massive aggregations (Palmer, et al., 2010). My next phase of work will explore and expand upon these connections, with the goal of more formally characterizing constellations of properties of thematic research collections that can contribute to generalizable principles of thematic gathering for scholarship within large digital libraries. For example, based on methods for identifying the presence of these properties, could we manually or automatically identify potential thematic research collections pooling in massive digital libraries or cultural heritage aggregations? Could we employ these properties as parameters to guide the development of digital library collections or *worksets* (Jett, 2015)? These questions continue a line of inquiry begun in Palmer, et al. (2010), which developed a method for identifying latent subject strengths in a large-scale aggregation of cultural heritage metadata according to attributes of item- and collection-level descriptions.

7.2.1. Flexibility, extensibility, and mobility

This section has reviewed the distinguishing purposes of different kinds of collections, but one purpose that all thematic research collections have in common is to serve as a platform for research on a theme. As a concluding reflection, I come back to a defining feature of thematic research collections, which seeps into every variety of collection purpose and design, and which fundamentally distinguishes collections from other genres of digital scholarship: that they serve as platforms for research. The remainder of this section considers novel properties of thematic research collections that arise from this study, and which have significant potential to advance

their capacity as platforms for research: the flexibility, extensibility, and mobility of collections. Finally, the section considers the implications of this analysis of defining features of thematic research collections for two further communities that should have a stake in the future of this genre: the library linked open data community, and communities at work on research and development of enhanced publications and related systems.

How has our sense of collection-as-platform changed since the earliest definitions of thematic research collections (especially Palmer, 2004)? How can we continue to evolve these platforms to support the next generation of humanities scholarship? Open-endedness has been a definitive characteristic of thematic research collections from the beginning; it is central to the accounts of Unsworth (2000) and Palmer (2004), and called out distinctly in the typology of Thomas (2016), as described above. Collections, unlike most published products of scholarship, remain indefinitely open to change and growth. Collections are flexible and extensible: beyond adding items, they can grow and change in how they support research. The ability to dynamically support research is central to a collection's role as platform.

My use of "platform" is meant to encompass several senses of the word: something that has been put in place to promote, to enable, and to make visible; a level space for opportunity; a layer upon which things may stand or be built. Thematic research collections function as purpose-built structures designed to support research and, in some cases, improvisation, projection, and performance (Nowvieskie, 2016). A collection as a platform must be more than a content store, as Palmer (2004) recognizes. It must also bear infrastructure, affordances, or access-provisions that lend themselves to generating new work. Here we dwell on the notion of generativity introduced in Chapter 4, but consider specific ways in which collections can help to generate meaning, new lines of inquiry, and collaboration. Palmer (2004) described how collections serve as platforms for interdisciplinary inquiry by gathering together diverse content to incorporate the interests of diverse intellectual communities. As they have evolved, collections have come to function as platforms in more and different ways, by encouraging different kinds of interactivity with the collection and its contents, and by building flexibility, extensibility, and mobility into their architectures.

Collections are *interactive* in different ways, and to different ends, ranging from facilitated interactivity to enabling potential reuse. Some collections create interactivity through built features, which afford dynamic but constrained modes of encountering and using items. For

example, the Shelley-Godwin Archive allows readers to view manuscripts in different orderings and with different encoded aspects illuminated. The Vault at Pfaff's allows users to navigate a multiplicity of different kinds of carefully encoded relationships among works and people. O Say Can You See allows users to traverse an expansive network of related people, families, and cases. Some collections facilitate interactivity in a social sense – not only interactivity with the collection itself, but with other users of the collection – turning the collection into a hub for collaboration or discourse, for example by adding forums and facilities for annotation, commenting, or feedback mechanisms. Other collections enable interactivity not by building specific uses or functions into the site, but by removing barriers to unanticipated use of collection contents and code, e.g., by opening their underlying data to use. This is the distinction between facilitated interactivity and enabling potential.

Collections enable unanticipated interaction with collections and their sources by designing architectures and data representations that are flexible, extensible, and mobile. These characteristics of collections constituted an unexpected and prominent theme emerging from the interviews. They offer new directions of potential growth for thematic research collections as platforms. While prior accounts of collections-as-platforms revolved around pulling things together into laboratory-like spaces with affordances, new models of “platformhood” may combine aggregation and affordance with more outward-looking features: features that allow sources within collections to be moved, interlinked, remixed, and repurposed outside of their original contexts. This kind of platform is enabled by purposive data modeling combined with multiple points of access and modes of acquisition.

Flexibility refers to the quality of being adaptable, versatile, and responsive to a variety of potential needs and interests. It is the quality of being available and accommodating to multiple user intentions. It is embodied by, for example, O Say Can You See's multiple points of access to its collection: through narrative stories, through legal cases, through individuals, through families as different navigational constructs. Shelley-Godwin manifests flexibility by allowing a general reader to see manuscript page images alongside a reader-friendly text; but also allowing scholars to see the TEI-XML encoding, or visit the archive for bulk access to the data. Flexibility asks: Do the data models at play – how items are described and represented – accommodate a variety of modes of discovery and access? One participant described how one prominent thematic research collection, the Rossetti Archive, conducted early experiments with flexibility, by trying to build in

the capacity for “remixing” contents according to users’ interests: “what would an interface look like if the primary scholar were an art historian rather than a literary scholar?” (P7). That experiment produced a prototype exhibit-builder, an affordance for allowing scholars to create tailored views and subsets of collections. Exhibit-building crosses the notion of flexibility for diverse potential audience with the notion of extensibility.

Closely related to flexibility, *extensibility* considers the extent to which a collection can be built upon and extended into new functions, new purposes, and new roles. This can happen either within the environment of the collection or outside of it. Is the collection self-contained, or does it invite new interdependencies? Is it open to inter-linking? Are there opportunities for co-creation? Does the architecture of the collection presume limited kinds of use, or does the collection consider and make space for unanticipated, even unanticipated-able uses? The “participatory platform” that the Shelley-Godwin Archive aspires to is an example of how a collection may afford extensibility within its environment. Shelley-Godwin hopes to enable users to build upon its collection by employing the Shared Canvas/IIIF data model, which opens the manuscript encodings to user annotations and integration with linked data. Collections can be extensible by being amenable to linking with other collections and data sources – both in the sense that their data models will support unique identification and interlinking, and in the sense that the architecture of the collection affords access to the data at that level.

Mobility refers to how collections can make themselves and their content movable or portable, open to adoption into new infrastructures, new institutions, and new contexts. This is not a novel concern for collections: NINES was meant to realize the aggregative power of thematic research collections by mobilizing collections as linked data in RDF. Mobility may be conceived as the ultimate expression of flexibility and extensibility. It can be facilitated alongside the thematic research collection as a performative site. A collection can enable different kinds of use and interaction, while also letting people take its underlying sources and value-added, interpretative data – such as encoded texts or encoded relationships between data components – elsewhere, into their own environments, for analysis, remodeling and reconstruction. Collections are mobilized in numerous ways: by the implementation of standards, APIs for direct access to data, open licensing, enabling bulk download of components, and thorough documentation of sources and their provenance. Not only was mobility described as of potential value for users, but

it was suggested as a long-term sustainability strategy: a collection that is mobile is self-contained, standards-compliant, and migration-ready, and therefore a more sustainable collection.

Scholarly collections gather and provide access to diverse sources (both primary and secondary), and build layers of interpretation and activity support on top of those sources. These are their foundational purposes. Both aspects – the sources themselves, and the layers built upon them – are valuable, and potentially share-able and reusable, contributions to scholarship. Their value for interpretation and reuse will depend on their own recognition and documentation of their “manufactured quality” (Flanders, 2014), of the decisions that went into their creation: How were sources selected? Why were the data modeled a certain way? What was foregrounded and omitted by transcribers and encoders, or by the extractive algorithms that produced a certain data set from a set of archival documents? What decisions and interpretations were made and enacted upon the sources, even in the course of their basic representation on the web?⁶¹ This level of documentation is something that most collections have struggled to produce.

Beyond these foundational purposes, collections can serve as generative platforms by facilitating experimentation and collaboration, and by producing new lines of inquiry and original evidence. They become platforms for what Nowvskie called “second-generation, born-digital scholarship.” Their generativity depends on how flexible, extensible, and mobile they are built to be. Flexibility and extensibility are implemented through the considerate selection and implementation of data models. Extensibility and mobility rely on the independence of underlying sources from the interpretive interfaces and affordances built on top of them. The goal of creating flexible, extensible, and mobile collections is to enhance the value that collections already add to sources by gathering them into rich, supportive contexts for research.

Linked open data (LOD) models, and systems built upon those models, appear to hold the most potential for undergirding collections that are flexible, extensible, and mobile. As we saw in the content analysis of collections, some collections are already adopting linked data standards for representation and description of primary sources and derived data. In recognition of the potential for LOD to enhance digital humanities research, the NINES aggregation years ago implemented an LOD-compatible model (based on the Resource Description Framework) for representing metadata aggregated from thematic research collections. That model has been adopted by the

⁶¹ This recalls Thomer’s (2017) assertion that scientific data reuse depends on researchers’ assessments of the (sometimes implicit) decisions and interpretations that have gone into the production of a data set.

organization that now subsumes NINES, the ARC consortium.⁶² However, most collections require substantial efforts to translate extant descriptive metadata about primary sources into the ARC RDF model. New systems for digital publishing are more commonly adopting or facilitating the use of linked data standards for resource representation and description, hopefully obviating some of this translational work for interconnecting digital humanities collections. For example, Omeka-S allows users to collect and publish items with linked open data descriptions.⁶³ While the momentum toward widespread adoption of LOD promises to improve access to and interoperability among digital collections, most thematic research collections do not take advantage of LOD for representing items or connecting to other resources. Future work will assess the potential applications of LOD for thematic research collections, with implications for LOD applications in contexts beyond digital humanities scholarship, including for library collections, and in enhanced publication data models.

7.2.2. Implications for libraries and linked open data

The kinds of thematic research collections identified in this study, together with the goals of flexibility, extensibility, and mobility for collection development, suggest new directions for exploiting humanities evidence scattered across the Web and in library, museum, and archival collections. Libraries, archives, and museums have recognized the potential benefits of LOD for representing digital collections for many years (Baker et al., 2011). Yet progress toward widespread implementation has been halting. Some libraries have begun transitioning technical services workflows into systems based on linked data standards. There are diffuse efforts to enrich digital special collections with semantic metadata to make connections to other collections and resources. As implementation spreads, there is a need for more empirical research on how LOD is used in scholarly research processes, and particularly how it may improve various uses of different kinds of digital collections. In future work, I aim to investigate scholarly use of linked open data in the context of thematic research collections; the implications of scholarly use for cultural collections in other contexts; and how LOD may be deployed to connect thematic research collections with one another, and with digitized cultural collections at libraries, archives, and museums, into networked platforms for research.

⁶² <http://idhmcmain.tamu.edu/arcgrant/about/aggregation/>

⁶³ <https://omeka.org/s/>

7.2.3. Implications for enhanced publications research and development

The representation of enhanced publications and complex research objects depends on data models that are readily expressed as and exploit the capacities of LOD. Thematic research collections are not fully served by extant descriptive and LOD standards for conventional, cultural collections, such as the Europeana Data Model. Thematic research collections stand to benefit from the extensible standards that have been developed for enhanced publications, which will need to be extended to accommodate properties that are essential to thematic research collections but not present in enhanced publications, such as variable ideals of *completeness*. The next phase of work will assess how enhanced publication data models and extant ontologies may be extended to accommodate the attributes of collections identified in this study, likely drawing on extant collection LOD models, such as the Europeana Data Model, and domain-specific vocabularies used in representing cultural heritage and humanities data sets. As this study has shown that common preservation streams for research data and digital objects in libraries struggle to accommodate the complexities of thematic research collections, repositories designed for enhanced publications may offer a way toward long-term management of fuller representations of thematic collections. The next phase of work, on extending data models to accommodate thematic research collections, also aims to identify ways in which aspects of enhanced publication management systems might extend the capacity for systemized management of complex digital scholarship.

7.3. CONCLUSIONS AND FUTURE WORK

This study suggests that the next generation of thematic research collections may serve as platforms that enable and facilitate research beyond their own boundaries. One possible and appealing future for the genre is as a diverse economy of moveable, flexible, repurposed and reusable collections of heterogeneous kinds of humanistic evidence, readily linkable to and embeddable in publications of born-digital interpretive scholarship. Realizing this potential will require us to find ways to accommodate collections within systems of scholarly communication, and especially established systems of scholarly credit and reputation-building. In turn, thematic research collections have opportunities to promote their own ends by interlinking with one another and encouraging users to build between and upon different collections.

This chapter has set out several directions in which my future work will build on the outcomes of this study. All are bent toward understanding how we may support the development of digital collections as contexts for research that are both tightly defined around scholarly interests, and at the same time open-ended, flexible, extensible, and mobile. In summary, these directions include:

- Scholarly use and reuse of thematic research collections, including for creating new, born-digital interpretive scholarship;
- The implications of defining features of thematic research collections, including purpose and completeness, for collections in other contexts, e.g., within aggregations or large-scale digital libraries;
- The sufficiency of collection-level descriptive schemas for describing thematic research collections in the context of different kinds of scholarly use;
- The challenges and opportunities for representing thematic research collections using linked open data, and implications for other kinds of cultural collections; and
- Extending data models for the representation of other genres of digital publication and collection to thematic research collections.

Researchers across disciplines increasingly seek to collect and share evidence in ways that meaningfully and reliably connect usable and repurpose-able data to essential context and interpretation, e.g., through hybrid publications, linked data, and complex research objects. Libraries have much at stake in the evolution of digital scholarship, and in thematic research collections specifically. The library's ability to serve its basic missions – of preserving the record of scholarship, advancing the research and teaching missions of universities, and serving broader public communities – will depend on how it rises to meet the demands of new scholarly communication practices and products. This study lays a foundation of understanding of thematic research collections as an exemplar of how digital scholarship in the humanities continues to evolve, and the implications for libraries as institutions charged with supporting and stewarding our scholarly and cultural records.

REFERENCES

- Acord, S. K., & Harley, D. (2013). Credit, time, and personality: The human challenges to sharing scholarly work using Web 2.0. *New Media & Society Theme Issue: "Scholarly Communication: Changes, Challenges & Initiatives,"* 15(3), 379–397.
- Adema, J., & Schmidt, B. (2010). From Service Providers to Content Producers: New Opportunities for Libraries in Collaborative Open Access Book Publishing. *New Review of Academic Librarianship*, 16(sup1), 28–43. <http://doi.org/10.1080/13614533.2010.509542>
- AHA Ad Hoc Committee on the Evaluation of Digital Scholarship by Historians. (2015). *Guidelines for the Evaluation of Digital Scholarship in History*. American Historical Association. Retrieved from <http://historians.org/teaching-and-learning/digital-history-resources/evaluation-of-digital-scholarship-in-history/guidelines-for-the-evaluation-of-digital-scholarship-in-history>
- Aimeur, E., Brassard, G., & Paquet, S. (2005). Personal knowledge publishing: fostering interdisciplinary communication. *IEEE Intelligent Systems*, 20(2), 46–53. <http://doi.org/10.1109/MIS.2005.34>
- Almas, B. (2017). Perseids: Experimenting with Infrastructure for Creating and Sharing Research Data in the Digital Humanities. *Data Science Journal*, 16(0). <https://doi.org/10.5334/dsj-2017-019>
- Almas, B., Berti, M., Choudhury, S., Dubin, D., Senseney, M., & Wickett, K. M. (2013, September). Representing Humanities Research Data Using Complementary Provenance Models. In *Poster presented at Building Global Partnerships—RDA Second Plenary Meeting, Washington, DC*.
- Alonso, C. J., Davidson, C. N., Unsworth, J. M., & Withey, L. (2003). *Crises and opportunities: The futures of scholarly publishing* (ACLS Occasional Paper No. 57). American Council of Learned Societies. Retrieved from https://www.acls.org/uploadedFiles/Publications/OP/57_Crises_and_Opportunities.pdf

- American Council of Learned Societies. (2006). *Our cultural commonwealth: The report of the American Council of Learned Societies Commission on cyberinfrastructure for the humanities and social sciences*. New York: American Council of Learned Societies.
Retrieved from
http://www.acls.org/uploadedFiles/Publications/Programs/Our_Cultural_Commonwealth.pdf
- American Historical Review. (2014). Call for submissions for the AHR Prize for best digital article. Retrieved August 28, 2017, from
http://www.indiana.edu/~ahrweb/Digital_Article_Prize.pdf
- Anderson, S., & McPherson, T. (2011). Engaging Digital Scholarship: Thoughts on Evaluating Multimedia Scholarship. *Profession*, 2011(1), 136–151.
<https://doi.org/10.1632/prof.2011.2011.1.136>
- ARC. (n.d.). Scholarly Peer Review. Retrieved August 28, 2017, from
<http://idhmcmain.tamu.edu/arcgrant/about/peer-review/>
- Assante, M., Candela, L., Castelli, D., Manghi, P., & Pagano, P. (2015). Science 2.0 Repositories: Time for a Change in Scholarly Communication. *D-Lib Magazine*, 21(1/2).
<https://doi.org/10.1045/january2015-assante>
- Attfield, S., & Dowell, J. (2003). Information seeking and use by newspaper journalists. *Journal of Documentation*, 59(2), 187–204. <http://doi.org/10.1108/00220410310463860>
- Baker, T., Bermès, E., Coyle, K., Dunsire, G., Isaac, A., Murray, P., ... Zeng, M. (2011). *Library Linked Data Incubator Group Final Report*. W3C Incubator Group Report. Retrieved from
<https://www.w3.org/2005/Incubator/lld/XGR-lld-20111025/>
- Ball, C. E. (2012). Assessing scholarly multimedia: A rhetorical Genre Studies approach. *Technical Communication Quarterly*, 21(1), 61–77.
<https://doi.org/10.1080/10572252.2012.626390>
- Bardi, A., & Manghi, P. (2014). Enhanced Publications: Data Models and Information Systems. *LIBER Quarterly*, 23(4). <https://doi.org/10.18352/lq.8445>

- Bardi, A., & Manghi, P. (2015). Enhanced Publication Management Systems: A systemic approach towards modern scientific communication. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1051–1052). New York, NY, USA: ACM. <https://doi.org/10.1145/2740908.2742026>
- Bates, D., Nelson, J., Roueché, C., Winters, J., & Wright, C. (2006). *Peer review and evaluation of digital resources for the arts and humanities: Final report and recommendations*. University of London: Institute of Historical Research. Retrieved from <http://www.history.ac.uk/projects/digital/peer-review>
- Beagrie, N. (2005). Plenty of room at the bottom? Personal digital libraries and collections. *D-Lib Magazine*, 11(6). Retrieved from <http://www.dlib.org/dlib/june05/beagrie/06beagrie.html>
- Beaudoin, J. E. (2012). Context and Its Role in the Digital Preservation of Cultural Objects. *D-Lib Magazine*, 18(11/12). <http://doi.org/10.1045/november2012-beaudoin1>
- Bechhofer, S., Roure, D. D., Gamble, M., Goble, C., & Buchan, I. (2010). Research Objects: Towards Exchange and Reuse of Digital Knowledge. *Nature Precedings*, (713). <https://doi.org/10.1038/npre.2010.4626.1>
- Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J., ... Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, 29(2), 599–611. <https://doi.org/10.1016/j.future.2011.08.004>
- Bell, P. (2000). Content analysis of visual images. In T. Van Leeuwen & C. Jewitt (Eds.), *The Handbook of Visual Analysis*. SAGE Publications.
- Blackburn, S. (2008). Fundamentum divisionis - Oxford Reference. In *The Oxford Dictionary of Philosophy (currently online version)* (2nd ed.). Oxford University Press. Retrieved from <http://www.oxfordreference.com/view/10.1093/acref/9780199541430.001.0001/acref-9780199541430-e-1329>
- Bloch, M. (1954). *The Historian's Craft*. Manchester: Manchester University Press. Retrieved from <http://catalog.hathitrust.org/api/volumes/oclc/22841856.html>

- Bonn, M., & Furlough, M. (Eds.). (2015). *Getting the word out: Academic libraries as scholarly publishers*. ACRL. Retrieved from <https://www.alastore.ala.org/detail.aspx?ID=11378>
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059–1078.
<https://doi.org/10.1002/asi.22634>
- Bourne, P. E., Clark, T., Dale, R., de Waard, A., Herman, I., Hovy, E., & Shotton, D. (2011). *FORCE11 Manifesto: Improving Future Research Communication and e-Scholarship*. Retrieved from <https://www.force11.org/about/manifesto>
- Brantley, S., Bruns, T., & Duffin, K. (2015). Leveraging OA, the IR, and Cross-Department Collaboration for Sustainability: Ensuring Library Centrality in the Scholarly Communication Discourse on Campus. *ACRL Proceedings 2015*. Retrieved from https://works.bepress.com/todd_bruns/55/
- Breure, L., Voorbij, H., & Hoogerwerf, M. (2011). Rich Internet Publications: “Show What You Tell.” *Journal of Digital Information*, 12(1). Retrieved from <https://journals.tdl.org/jodi/index.php/jodi/article/view/1606>
- Brockman, W. S., Neumann, L., Palmer, C. L., & Tidline, T. J. (2001). *Scholarly work in the humanities and the evolving information environment* (No. 104). Washington, D.C.: Digital Library Federation, Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub104/pub104.pdf>
- Brogan, M. (2006). *Contexts and contributions: Building the distributed library*. Digital Library Federation/Council on Library and Information resources.
- Brown, L., Griffiths, R., & Rascoff, M. (2007). *University publishing in a digital age*. New York: ITHAKA. Retrieved from <http://www.sr.ithaka.org/sites/default/files/reports/4.13.1.pdf>
- Brown, S., & Simpson, J. (2014). The changing culture of humanities scholarship: Iteration, recursion, and versions in scholarly collaboration environments. *Scholarly and Research Communication*, 5(4). Retrieved from <http://src-online.ca/src/index.php/src/article/view/191/354>

- Buchanan, G., Bainbridge, D., Don, K. J., & Witten, I. H. (2005). A new framework for building digital library collections. *Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '05)*, 23–31. <http://doi.org/10.1145/1065385.1065392>
- Bulger, M. E., Meyer, E. T., De la Flor, G., Terras, M., Wyatt, S., Jirotko, M., ... Madsen, C. (2011). *Reinventing Research? Information Practices in the Humanities*. Research Information Network. Retrieved from <http://dx.doi.org/10.2139/ssrn.1859267>
- Calhoun, K. (2011). The changing nature of the catalog and its integration with other discovery tools. In J. McIntosh (Ed.), *Cataloging and Indexing: Challenges and Solutions*. CRC Press.
- Caprio, M. J. (2015). Re-Engineering Relationships with Faculty and Students: A Social Contract for Digital Scholarship. In B. L. Eden (Ed.), *Creating Research Infrastructures in 21st-Century Academic Libraries: Conceiving, Funding, and Building New Facilities and Staff*. Rowman & Littlefield. Retrieved from <https://rowman.com/ISBN/9781442252400/Creating-Research-Infrastructures-in-the-21st-Century-Academic-Library-Conceiving-Funding-and-Building-New-Facilities-and-Staff>
- Center for Digital Research in the Humanities. (n.d.). Best Practices for Digital Humanities. Retrieved April 5, 2016, from http://unlcms.unl.edu/cas/center-for-digital-research-in-the-humanities/articles/best_practices
- Ciula, A., & Lopez, T. (2009). Reflecting on a dual publication: Henry III Fine Rolls print and web. *Literary and Linguistic Computing*, 24(2), 129–141. <http://doi.org/10.1093/lc/fqp007>
- Clement, T., Hagenmaier, W., & Knies, J. L. (2013). Toward a Notion of the Archive of the Future: Impressions of Practice by Librarians, Archivists, and Digital Humanities Scholars. *The Library Quarterly*, 83(2), 112–130. <http://doi.org/10.1086/669550>
- Cohen, D. J. (2013). The social contract of scholarly publishing. In *Debates in the Digital Humanities* (Open Access Edition). Retrieved from <http://dhdebates.gc.cuny.edu/debates/text/27>
- Cohen, D. J., & Fragaszy Troyano, J. (2012). Closing the evaluation gap [Introduction to issue of Journal of the Digital Humanities]. *Journal of Digital Humanities*, 1(4). Retrieved from <http://journalofdigitalhumanities.org/1-4/closing-the-evaluation-gap/>

- Cohen, D. J., & Rosenzweig, R. (2005). Preserving Digital History. In *Digital history: A guide to gathering, preserving, and presenting the past on the Web*. University of Pennsylvania Press. Retrieved from <http://chnm.gmu.edu/digitalhistory/preserving/index.php>
- Coletta, C. D. (2011). Guidelines for promotion and tenure committees in judging digital work. Retrieved from <http://institutes.nines.org/docs/2011-documents/guidelines-for-promotion-and-tenure-committees-in-judging-digital-work/>
- Collier, D., Laporte, J., & Seawright, J. (2008). Typologies: Forming Concepts and Creating Categorical Variables. Retrieved from <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199286546.001.0001/oxfordhb-9780199286546-e-7>
- Corrall, S., & Roberts, A. (2012). Information Resource Development and “Collection” in the Digital Age: Conceptual Frameworks and New Definitions for the Network World. *Libraries in the Digital Age (LIDA) Proceedings*, 12(0). Retrieved from <http://ozk.unizd.hr/proceedings/index.php/lida/article/view/62>
- Corrall, S., & Roberts, A. (2014). Collection as thing, process, and access: Two proposed models. Presented at the Digital Collection Contexts: Intellectual and Organizational Functions at Scale, Berlin, Germany: University of Pittsburgh. Retrieved from <http://d-scholarship.pitt.edu/22015/>
- Council on Library and Information Resources. (2010). *The idea of order: transforming research collections for 21st century scholarship* (No. 147). Washington, D.C.: CLIR. Retrieved from <http://www.clir.org/pubs/abstract/reports/pub147>
- Courant, P., & Jones, E. (2015). Scholarly Publishing as an Economic Public Good. In M. Bonn & M. Furlough (Eds.), *Getting the Word Out: Academic Libraries as Scholarly Publishers* (pp. 17–42). ACRL.
- Cragin, M. H., Heidorn, P. B., Palmer, C. L., & Smith, L. C. (2007). An Educational Program on Data Curation. In *Science and Technology Section*. Washington, D.C.

- Crow, R. (2009). *Campus-based publishing partnerships: A guide to critical issues*. Washington, D.C.: SPARC. Retrieved from http://www.sparc.arl.org/sites/default/files/pub_partnerships_v1.pdf
- Davidson, S. B., & Freire, J. (2008). Provenance and scientific workflows: challenges and opportunities. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data* (pp. 1345-1350). ACM.
- Dempsey, L. (2006). The (digital) library environment: Ten years after. *Ariadne*, 46.
- De Roure, D. (2014). The future of scholarly communications. *Insights*, 27(3). <https://doi.org/10.1629/2048-7754.171>
- DHCommons. (n.d.). Review guidelines. Retrieved August 28, 2017, from <http://dhcommons.org/journal/review-guidelines>
- Digital Humanities Working Group, University of Florida. (n.d.). DH in Tenure & Promotion. Retrieved April 5, 2016, from <http://digitalhumanities.group.ufl.edu/dh-uf/tenure-promotion/>
- Doerr, M. (2014). Unity criteria. *Digital Collection Contexts: iConference 2014 Workshop Report*. Retrieved from <https://www.ideals.illinois.edu/bitstream/handle/2142/73359/DCCWorkshopReport.pdf?sequence=2>
- Drucker, J. (2009). Blind spots. *Chronicle of Higher Education*, 55(30).
- Duff, W. M., & Johnson, C. A. (2002). Accidentally found on purpose: Information-seeking behaviors of historians in archives. *Library Quarterly*, 72(4), 472–496.
- Duranti, L. (2005). The long-term preservation of accurate and authentic digital data: the INTERPARES project. *Data Science Journal*, 4, 106–118. <https://doi.org/10.2481/dsj.4.106>
- Edmond, J. (2016). Will historians ever have big data? Theoretical and infrastructural perspectives. In *Computational History and Data-Driven Humanities: Second IFIP WG 12.7 International Workshop, CHDDH 2016, Dublin, Ireland, May 25, 2016, Revised Selected Papers 2* (pp. 91-105). Springer International Publishing.

- Efron, M., Organisciak, P., & Fenlon, K. (2011). Building topic models in a federated digital library through selective document exclusion. In *Proceedings of the Annual Meeting of the American Society for Information Science and Technology*.
- Feng, L., Jeusfeld, M. A., & Hoppenbrouwers, J. (2004). Beyond information searching and browsing: acquiring knowledge from digital libraries. *Information Processing and Management*, 41, 97–120.
- Fenlon, K. (2017). Thematic research collections: Libraries and the evolution of alternative digital publishing in the humanities. *Library Trends*, 65(4), 523–539.
- Fenlon, K., Senseney, M., Green, H., Bhattacharyya, S., Willis, C., & Downie, J. S. (2014). Scholar-built collections: A study of user requirements for research in large-scale digital libraries. In *Proceedings of the ASIS&T Annual Meeting*. Seattle, WA.
- Fitzpatrick, K. (2011). Peer Review. In *Planned Obsolescence: Publishing, Technology, and the Future of the Academy*. New York: NYU Press.
- Fitzpatrick, K. (2015). Peer Review. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 439–448). John Wiley & Sons, Ltd. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781118680605.ch30/summary>
- Flanders, J. (2014). Rethinking Collections. In P. L. Arthur & K. Bode (Eds.), *Advancing Digital Humanities* (pp. 163–174). Palgrave Macmillan UK. Retrieved from http://link.springer.com/chapter/10.1057/9781137337016_11
- Flanders, J., & Muñoz, T. (2012). An Introduction to Humanities Data Curation. In *DH Curation: A Community Resource Guide to Data Curation for the Digital Humanities*. Champaign, IL. Retrieved from <http://guide.dhcuration.org/intro/>
- Fleming-May, R. A. (2011). What Is Library Use? Facets of Concept and a Typology of Its Application in the Literature of Library and Information Science. *Library Quarterly*, 81(3), 297–320.
- Fortier, R., & James, H. (2015). Becoming the Gothic Archive: From digital collection to Digital Humanities. In K. Sacco (Ed.), *Supporting Digital Humanities for Knowledge Acquisition in Modern Libraries*. IGI Global.

- Goddard, L., & Walde, C. (2017). Negotiating sustainability: The grant services “menu” at UVic Libraries. Presented at the Digital Humanities, ADHO. Retrieved from <https://dh2017.adho.org/abstracts/231/231.pdf>
- Hahn, K. L. (2008). *Research library publishing services: New options for university publishing*. Washington, D.C.: Association of Research Libraries.
- Hahn, K., Lowry, C., Lynch, C., Shulenberger, D., & Vaughn, J. (2009). *The university’s role in the dissemination of research and scholarship -- A call to action*. AAU, ARL, CNI, NASULGC (APLU). Retrieved from <http://files.eric.ed.gov/fulltext/ED511357.pdf>
- Harley, D. (2013). Scholarly Communication: Cultural Contexts, Evolving Models. *Science*, 342(6154), 80–82. <http://doi.org/10.1126/science.1243622>
- Harley, D., Acord, S. K., & Earl-Novell, S. (2010). *Peer review in academic promotion and publishing: Its meaning, locus and future*. Berkeley, CA: Center for Studies in Higher Education.
- Harley, D., Acord, S. K., Earl-Novell, S., Lawrence, S., & Judson King, C. (2010). *Assessing the Future Landscapes of Scholarly Communication: an Exploration of Faculty Values and Needs in Seven Disciplines*. Berkeley, CA: Center for Studies in Higher Education.
- Hedstrom, M., & Lee, C. A. (2002). Significant properties of digital objects: Definitions, applications, implications. *Proceedings of the DLM-Forum*, 218–227.
- Henny, U., Neuber, F., & IDE. (2017). *Criteria for reviewing digital text collections, version 1.0*. Institut für Dokumentologie und Editorik (IDE). Retrieved from <http://www.i-d-e.de/publikationen/weitereschriften/criteria-text-collections-version-1-0/>
- Heyvaert, P., De Nies, T., Van Herwegen, J., Vander Sande, M., Verborgh, R., De Neve, W., ... Van de Walle, R. (2015). Using EPUB3 and the Open Web Platform for enhanced presentation and machine-understandable metadata for digital comics. In *New Avenues for Electronic Publishing in the Age of Infinite Collections and Citizen Science: Scale, Openness and Trust - Proceedings of the 19th International Conference on Electronic Publishing*. IOS Press Ebooks. <https://doi.org/10.3233/978-1-61499-562-3-37>

- Hill, L. L., Janee, G., Dolin, R., Frew, J., & Larsgaard, M. (1999). Collection Metadata Solutions for Digital Library Applications. *Journal of the American Society for Information Science*, 50(13).
- Hoekstra, R., Meroño-Peñuela, A., Dentler, K., Rijpma, A., Zijdeman, R., & Zandhuis, I. (2016). An ecosystem for linked humanities data. In *The Semantic Web* (pp. 425–440). Springer, Cham. https://doi.org/10.1007/978-3-319-47602-5_54
- Horava, T. (2011). Challenges and Possibilities for Collection Management in a Digital Age. *Library Resources & Technical Services*, 54(3), 142–152.
- Hou, C.-Y., Thompson, C. A., & Palmer, C. L. (2014). Profiling open digital repositories in the atmospheric and climate sciences: An initial survey. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1–4.
<http://doi.org/10.1002/meet.2014.14505101121>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <http://doi.org/10.1177/1049732305276687>
- Jankowski, N. W. (2011). *Enhancing Scholarly Publishing in the Humanities and Social Sciences: Innovation Through Hybrid Forms of Publication* (SSRN Scholarly Paper No. ID 1929687). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=1929687>
- Jankowski, N. W. (2011). Why Enhanced Publications? Retrieved November 7, 2017, from <http://ep-books.ehumanities.nl/why-enhanced-publications>
- Jankowski, N. W., Scharnhorst, A., Tatum, C., & Tatum, Z. (2013). Enhancing scholarly publications: Developing hybrid monographs in the humanities and social sciences. *Scholarly and Research Communication*, 4(1). Retrieved from <http://src-online.ca/index.php/src/article/view/40>
- Jewell, A. (2009). Digital Editions: Scholarly Tradition in an Avant-Garde Medium. *Documentary Editing*, 30(3-4), 28–35.
- Johnston, P., & Robinson, B. (2002). *Collections and collection description* (Briefing Paper No. 1). UKOLN.

- Kakali, C., Lourdi, I., Stasinopoulou, T., Bountouri, L., Papatheodorou, C., Doerr, M., & Gergatsoulis, M. (2007). Integrating Dublin Core Metadata for Cultural Heritage Collections Using Ontologies. *International Conference on Dublin Core and Metadata Applications, 0(0)*, pp. 128–139.
- Kakar, A. S. (2016). A User-Centric Typology of Information System Requirements. *Journal of Organizational & End User Computing*, 28(1), 32–55.
<http://doi.org/10.4018/JOEUC.2016010103>
- Kang, K. C., Cohen, S. G., Hess, J. A., Novak, W. E., & Peterson, A. S. (1990). *Feature-Oriented Domain Analysis (FODA) Feasibility Study*. Retrieved from
<http://www.engr.sjsu.edu/fayad/current.courses/cmpe202-fall2013/docs/CmpE202-SE-Link-Part-Two-Fall2013/11-Domain%20Analysis/90tr021.pdf>
- Kling, R. (2005). The internet and unrefereed scholarly publishing. *Annual Review of Information Science and Technology*, 38(1), 591–631.
<http://doi.org/10.1002/aris.1440380113>
- Kluge, S. (2000). Empirically Grounded Construction of Types and Typologies in Qualitative Social Research. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(1). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1124>
- Koch, T. (2000). Quality-controlled subject gateways: definitions, typologies, empirical overview. *Online Information Review*, 24(1), 24–34.
<http://doi.org/10.1108/14684520010320040>
- Kuhn, V., Johnson, D. J., & Lopez, D. (2010). Speaking with students: Profiles in digital pedagogy. *Kairos*, 14(2). Retrieved from
<http://kairos.technorhetoric.net/14.2/interviews/kuhn/index.html>
- Lagoze, C., & Fielding, D. (1998). Defining collections in distributed digital libraries. *D-Lib Magazine*. Retrieved from <http://www.dlib.org/dlib/november98/lagoze/11lagoze.html>
- Lee, H. (2005). The Concept of Collection from the Users' Perspective. *Library Quarterly*, 75(1), 67–85.

- Lee, H. L. (2000). What is a Collection? *Journal of the American Society for Information Science*, 51(12), 1106–1113.
- Lefevre, J., & Huwe, T. K. (2013). Digital Publishing from the Library: A New Core Competency. *Journal of Web Librarianship*, 7(2), 190–214.
<http://doi.org/10.1080/19322909.2013.780519>
- Lippincott, S. (Ed.). (2015). Library Publishing Directory 2015. Library Publishing Coalition.
 Retrieved from
http://www.librarypublishing.org/sites/librarypublishing.org/files/documents/lpc_dir_2015lpd.pdf
- Lourdi, I., Papatheodorou, C., & Doerr, M. (2009). Semantic Integration of Collection Description: Combining CIDOC/CRM and Dublin Core Collections Application Profile. *D-Lib Magazine*, 15(7/8). <http://doi.org/10.1045/july2009-papatheodorou>
- Maistrovich, T. (2014). Typology of Electronic Libraries. *Slavic & East European Information Resources*, 15(4), 240–246. <http://doi.org/10.1080/15228886.2014.970915>
- Mandell, L. (2012). Promotion and Tenure for Digital Scholarship. *Journal of Digital Humanities*. 1(4). Retrieved from <http://journalofdigitalhumanities.org/1-4/promotion-and-tenure-for-digital-scholarship-by-laura-mandell/>
- Marcial, L. H., & Hemminger, B. M. (2010). Scientific data repositories on the Web: An initial survey. *Journal of the American Society for Information Science and Technology*, 61(10), 2029–2048. <http://doi.org/10.1002/asi.21339>
- Marradi, A. (1990). Classification, typology, taxonomy. *Quality and Quantity*, 24(2), 129–157.
<http://doi.org/10.1007/BF00209548>
- Mattern, S. C. (2012). Evaluating Multimodal Work, Revisited. *Journal of Digital Humanities* 1(4). Retrieved from <http://journalofdigitalhumanities.org/1-4/evaluating-multimodal-work-revisited-by-shannon-mattern/>
- Mayring, P. (2000). Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 1(2). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1089>

- McCormick, M. (2015). Toward new-model scholarly publishing: Uniting the skills of publishers and libraries. In M. Bonn & M. Furlough (Eds.), *Getting the word out: Academic libraries as scholarly publishers*. ACRL.
- McFall, L. M. (2015). Beyond the back room: The role of metadata and catalog librarians in digital humanities. In K. Sacco (Ed.), *Supporting Digital Humanities for Knowledge Acquisition in Modern Libraries*. IGI Global.
- Meiman, M. (2015). *Digitizing the nineteenth century: scholarly editing, interface design, and affordances for public engagement* (Department of English). University of Delaware. Retrieved from <http://udspace.udel.edu/handle/19716/17209>
- Modern Language Association. (2012). *Guidelines for Evaluating Work in Digital Humanities and Digital Media* / Modern Language Association. Modern Language Association. Retrieved from <https://www.mla.org/About-Us/Governance/Committees/Committee-Listings/Professional-Issues/Committee-on-Information-Technology/Guidelines-for-Evaluating-Work-in-Digital-Humanities-and-Digital-Media>
- Mueller, M. (2010). *Towards a digital carrel: A report about corpus query tools. Report from Mellon Workshop. Evanston, Illinois. Nov. 22-23.* Evanston, IL: Mellon Workshop at Northwestern University.
- Mullins, J., Murray-Rust, C., Ogburn, J., Crow, R., Ivins, O., Mower, A., ... Watkinson, C. (2012). Library Publishing Services: Strategies for Success: Final Research Report (March 2012). *Purdue University Press E-Books*, 24.
- Muñoz, T. (2013, May 30). Data curation as publishing for digital humanists. Retrieved from <http://www.trevormunoz.com/notebook/>
- Nowviskie, B. (2011). a skunk in the library. Retrieved September 6, 2016, from <http://nowviskie.org/2011/a-skunk-in-the-library/>
- Nowviskie, B. (2012). Evaluating Collaborative Digital Scholarship (or, Where Credit is Due). *Journal of Digital Humanities*. 1(4). Retrieved from <http://journalofdigitalhumanities.org/1-4/evaluating-collaborative-digital-scholarship-by-bethany-nowviskie/>

- OED (2017). “definitive, n.1”. *OED Online*. Oxford University Press. Retrieved from <https://en.oxforddictionaries.com/definition/definitive>
- Organization of American Historians. (2013). Digital History Reviews. Retrieved from <http://jah.oah.org/submit/digital-history-reviews/>
- Ortega, C. D. (2012). Conceptual and Procedural Grounding of Documentary Systems. *Knowledge Organization*, 39(3), 224–228.
- Padilla, T. (2016). Humanities data in the library: Integrity, form, access. *D-Lib*, 22(3/4). <https://doi.org/10.1045/march2016-padilla>
- Page, K., Lewis, D., & Weigl, D. (2017). Contextual interpretation of digital music notation. Presented at the Digital Humanities (DH2017), Montréal, Canada.
- Palmer, C. L. (2004). Thematic Research Collections. In *A Companion to Digital Humanities*. Blackwell Publishing. Retrieved from http://digitalhumanities.org:3030/companion/view?docId=blackwell/9781405103213/9781405103213.xml&chunk.id=ss1-4-5&toc.depth=1&toc.id=ss1-4-5&brand=9781405103213_brand
- Palmer, C. L. (2005). Scholarly work and the shaping of digital access. *Journal of the American Society for Information Science and Technology*, 56(11), 1140–1153. <http://doi.org/10.1002/asi.20204>
- Palmer, C. L., & Fenlon, K. (2017). Information research on interdisciplinarity. In R. Frodeman, J. T. Klein, & R. C. D. S. Pacheco (Eds.), *The Oxford Handbook of Interdisciplinarity* (Second Edition). Oxford, New York: Oxford University Press.
- Palmer, C. L., & Neumann, L. J. (2002). The information work of interdisciplinary humanities scholars: Exploration and translation. *The Library Quarterly*, 72(1), 85–117.
- Palmer, C. L., Teffeau, L. C., & Pirmann, C. M. (2009). *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development*. Dublin, OH: OCLC Research and Programs. Retrieved from <http://www.oclc.org/content/dam/research/publications/library/2009/2009-02.pdf>

- Palmer, C. L., Weber, N. M., & Cragin, M. H. (2011). The analytic potential of scientific data: Understanding re-use potential. *Proceedings of the American Society for Information Science & Technology*, 48(1).
- Palmer, C. L., Zavalina, O., & Fenlon, K. (2010). Beyond size and search: Building contextual mass in digital aggregations for scholarly use. In *Proceedings of the ASIS&T Annual Meeting*. Pittsburgh, PA. Retrieved from <http://hdl.handle.net/2142/18655>
- Patton, M. Q. (2002). *Qualitative Research and Evaluation Methods*. Thousand Oaks, CA: Sage.
- Pejšová, P., & Vaska, M. (2011). An Analysis of Current Grey Literature Document Typology. *Grey Journal (TGJ)*, 7(2), 72–80.
- Pe-Than, E. P. P., Goh, D. H.-L., & Lee, C. S. (2015). A typology of human computation games: an analysis and a review of current games. *Behaviour & Information Technology*, 34(8), 809–824. <http://doi.org/10.1080/0144929X.2013.862304>
- Presner, T. (2012, December 19). How to Evaluate Digital Scholarship. *Journal of Digital Humanities*. 1(4). Retrieved from <http://journalofdigitalhumanities.org/1-4/how-to-evaluate-digital-scholarship-by-todd-presner/>
- Price, K. M. (2009). Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name? *Digital Humanities Quarterly*, 3(3). Retrieved from <http://www.digitalhumanities.org/dhq/vol/3/3/000053/000053.html>
- Procter, R., Williams, R., Stewart, J., Poschen, M., Snee, H., Voss, A., & Asgari-Targhi, M. (2010). Adoption and use of Web 2.0 in scholarly communications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 368(1926), 4039–4056. <http://doi.org/10.1098/rsta.2010.0155>
- Ramsay, S. (2014). The hermeneutics of screwing around; or What you do with a million books. In K. Kee (Ed.), *Pastplay: Teaching and Learning History with Technology*. Ann Arbor, MI: Michigan Publishing. Retrieved from <http://dx.doi.org/10.3998/dh.12544152.0001.001>
- Renear, A. H., Wickett, K. M., Urban, R. J., Dubin, D., & Shreeves, S. L. (2008). Collection/Item Metadata Relationships. *International Conference on Dublin Core and Metadata Applications*, 0(0), 80–89.

- Roberts, A. (2014, February). *Conceptualising the library collection for the digital world: A case study of social enterprise* (Thesis). University of Sheffield. Retrieved from <http://etheses.whiterose.ac.uk/5186/>
- Rockwell, G. (2011). On the evaluation of digital media as scholarship. *Profession*, 11, 152–168.
- Rockwell, G. (2012). Short Guide To Evaluation Of Digital Work. *Journal of Digital Humanities*. 1(4). Retrieved from <http://journalofdigitalhumanities.org/1-4/short-guide-to-evaluation-of-digital-work-by-geoffrey-rockwell/>
- Rorissa, A. (2010). A comparative study of Flickr tags and index terms in a general image collection. *Journal of the American Society for Information Science and Technology*, 61(11), 2230–2242. <http://doi.org/10.1002/asi.21401>
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7(4), 573–605. [http://doi.org/10.1016/0010-0285\(75\)90024-9](http://doi.org/10.1016/0010-0285(75)90024-9)
- Rosenzweig, R. (2003). Scarcity or Abundance? Preserving the Past in a Digital Era. *The American Historical Review*, 108(3), 735–762. <https://doi.org/10.1086/529596>
- Rousi, A. M., Savolainen, R., & Vakkari, P. (2016). A typology of music information for studies on information seeking. *Journal of Documentation*, 72(2), 265–276. <http://doi.org/10.1108/JD-01-2015-0018>
- Sahle, P., Vogeler, G., & IDE. (2014). *Criteria for Reviewing Scholarly Digital Editions, version 1.1*. Institut für Dokumentologie und Editorik. Retrieved from <https://www.i-d-e.de/publikationen/weitereschriften/criteria-version-1-1/>
- Schöch, C. (2013). Big? Smart? Clean? Messy? Data in the Humanities. *Journal of Digital Humanities*, 2(3), 2–13.
- Schreibman, S., Roper, J. O., & Gueguen, G. (2008). Cross-collection Searching: A Pandora's Box or the Holy Grail? *Literary and Linguistic Computing*, 23(1), 13–25. <http://doi.org/10.1093/llc/fqm039>

- Schreier, M. (2013). Qualitative content analysis. In U. Flick (Ed.), *The SAGE Handbook of Qualitative Data Analysis*. London: SAGE Publications.
- Sigarchian, H. G., De Meester, B., De Nies, T., Verborgh, R., De Neve, W., Mannens, E., & Van De Walle, R. (2014). EPUB3 for integrated and customizable representation of a scientific publication and its associated resources. In *Proceedings of the 4th International Conference on Linked Science - Volume 1282* (pp. 1–11). Aachen, Germany, Germany: CEUR-WS.org. Retrieved from <http://dl.acm.org/citation.cfm?id=2878584.2878585>
- Sinn, D., & Soares, N. (2014). Historians' use of digital archival collections: The web, historical scholarship, and archival research. *Journal of the Association for Information Science and Technology*, 65(5). <http://doi.org/10.1002/asi.23091>
- Smit, E., Van Der Hoeven, J., & Giaretta, D. (2011). Avoiding a Digital Dark Age for data: why publishers should care about digital preservation. *Learned Publishing*, 24(1), 35–49. <https://doi.org/10.1087/20110107>
- Smithies, J. (2012, December 19). Evaluating Scholarly Digital Outputs: The Six Layers Approach. *Journal of Digital Humanities*. 1(4). Retrieved from <http://journalofdigitalhumanities.org/1-4/evaluating-scholarly-digital-outputs-by-james-smithies/>
- Stodden, V. C. (2010). Reproducible research: Addressing the need for data and code sharing in computational science. *Computing in Science & Engineering*, 12(5), 8–12. <https://doi.org/10.1109/MCSE.2010.113>
- Stvilia, B., & Jørgensen, C. (2009). User-generated collection-level metadata in an online photo-sharing system. *Library & Information Science Research*, 31(1), 54–65. <http://doi.org/10.1016/j.lisr.2008.06.006>
- Sukovic, S. (2008). Convergent flows: Humanities scholars and their interactions with electronic texts. *The Library*, 78(3). Retrieved from <http://ses.library.usyd.edu.au/bitstream/2123/3570/1/sukovic%20convergent%20flows.pdf>
- Sukovic, S. (2011). E-Texts in Research Projects in the Humanities. In A. Woodsworth & W. D. Penniman (Eds.), *Advances in Librarianship*. Bingley, UK: Emerald Group Publishing.

- Sundaram, F. (2016, March 21). Publishing Digital Scholarship: Perspectives from Stanford University Press [Digital Library Federation]. Retrieved from <https://www.diglib.org/archives/11455/>
- Sustaining Digital Scholarship. (2004). *SDS Final Report*. University of Virginia: Institute for Advanced Technology in the Humanities. Retrieved from https://dcs.library.virginia.edu/files/2012/05/SDS_FinalReport2003.pdf
- Svenonius, E. (2000). *The Intellectual Foundation of Information Organization*. MIT Press.
- Tennis, J. T. (2011). Is There a New Bibliography? *Cataloging & Classification Quarterly*, 49(2), 121–126. <http://doi.org/10.1080/01639374.2011.544020>
- Thibodeau, K. (2002). Overview of technological approaches to digital preservation and challenges in coming years. In *The State of Digital preservation: An International Perspective*. CLIR and the Library of Congress. <https://doi.org/10.1.1.89.3273>
- Thomas, W. G. (2015). The Promise of the Digital Humanities and the Contested Nature of Digital Scholarship. In S. Schreibman, R. Siemens, & J. Unsworth (Eds.), *A New Companion to Digital Humanities* (pp. 524–537). John Wiley & Sons, Ltd. Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9781118680605.ch36/summary>
- Thomas, W. G., & Ayers, E. L. (2003). An Overview: The Differences Slavery Made: A Close Analysis of Two American Communities. *The American Historical Review*, 108(5), 1299–1307. <https://doi.org/10.1086/529967>
- Thomer, A. K. (2017). *Site-based data curation: bridging data collection protocols and curatorial processes at scientifically significant sites*. University of Illinois at Urbana-Champaign. Retrieved from <http://hdl.handle.net/2142/98372>
- Tibbo, H. (2003). Primarily history in America: How U.S. historians search for primary materials at the dawn of the digital age. *The American Archivist*, 66(1), 9–50.
- Toms, E. G., & Flora, N. (2005). From physical to digital humanities library – designing the humanities scholar’s workbench. In R. Siemens & D. Moorman (Eds.), *Mind Technologies: Humanities Computing and the Canadian Academic Community* (pp. 91–115). Calgary: University of Calgary Press.

- Toms, E. G., & O'Brien, H. (2008). Understanding the information and communication technology needs of the e-humanist. *Journal of Documentation*, 64(1), 102–130. <http://doi.org/10.1108/00220410810844178>
- Tracy, D. G. (2016). Assessing Digital Humanities Tools: Use of Scalar at a Research University. *Portal: Libraries and the Academy*, 16(1), 165–191.
- Unsworth, J. (2000). Thematic Research Collections. Presented at the Modern Languages Association Annual Conference, Washington, D.C. Retrieved from <http://www.iath.virginia.edu/~jmu2m/MLA.00/>.
- Unsworth, J. (2003). The crisis in scholarly publishing in the humanities. *ARL: A Bimonthly Report on Research Library Issues and Actions from ARL, CNI, and SPARC*, 1–4.
- Van de Sompel, H., Payette, S., Erickson, J., Lagoze, C., & Warner, S. (2004). Rethinking Scholarly Communication: Building the System that Scholars Deserve. *D-Lib Magazine*, 10(9). <https://doi.org/10.1045/september2004-vandesompel>
- Vandegrift, M. (2012). What Is Digital Humanities and What's it Doing in the Library? *In the Library with the Lead Pipe*. Retrieved from <http://www.inthelibrarywiththeleadpipe.org/2012/dhandthelib/>
- Vandegrift, M., & Varner, S. (2013). Evolving in Common: Creating Mutually Supportive Relationships Between Libraries and the Digital Humanities. *Journal of Library Administration*, 53(1), 67–78. <http://doi.org/10.1080/01930826.2013.756699>
- Vanwynsberghe, H., Vanderlinde, R., Georges, A., & Verdegem, P. (2015). The librarian 2.0: Identifying a typology of librarians' social media literacy. *Journal of Librarianship & Information Science*, 47(4), 283–293. <http://doi.org/10.1177/0961000613520027>
- Varvel, V. E. J., & Thomer, A. (2011). *Google Digital Humanities Awards Recipient Interviews Report* (No. CIRSS Report No. HTRC1101). Champaign, IL: Center for Information Research in Science and Scholarship, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign. Retrieved from <http://hdl.handle.net/2142/29936>

- Verhaar, P. A. F. (2008). *Report on object models and functionalities* (No. D4.2). DRIVER, Digital Repository Infrastructure Vision for European Research II. Retrieved from https://openaccess.leidenuniv.nl/bitstream/handle/1887/16018/Report_on_Object_Models_and_Functionalities.pdf
- Waismann, F. (1951). Verifiability. In A. Flew (Ed.), *Logic and Language*. Retrieved from <http://www.ditext.com/waismann/verifiability.html>
- Warner, A. (2007). Constructing a tool for assessing scholarly webtexts. *Kairos*, 12(1). Retrieved from <http://kairos.technorhetoric.net/12.1/binder.html?topoi/warner/index.html>
- Warwick, C., Terras, M., Galina, I., Huntington, P., & Pappa, N. (2007). Evaluating digital humanities resources: The LAIRAH project checklist and the Internet Shakespeare Editions project. In *Proceedings ELPUB 2007*. Vienna, Austria. Retrieved from <http://dro.dur.ac.uk/15194/1/15194.pdf>
- Watkinson, C., Skinner, K., & Speer, J. (2012). *Library Publishing Coalition: A Proposal*. Retrieved from http://www.librarypublishing.org/sites/librarypublishing.org/files/documents/lpc_proposal_20120814.pdf
- Weller, M. (2011). *The Digital Scholar: How Technology is Transforming Scholarly Practice*. A&C Black. Retrieved from <https://www.bloomsburycollections.com/book/the-digital-scholar-how-technology-is-transforming-scholarly-practice/>
- White, M. D., & Marsh, E. E. (2006). Content Analysis: A Flexible Methodology. *Library Trends*, 55(1), 22–45. <http://doi.org/10.1353/lib.2006.0053>
- Wickett, K. M. (2012). *Collection/item metadata relationships*. University of Illinois at Urbana-Champaign, Champaign, IL. Retrieved from <http://hdl.handle.net/2142/42198>
- Wickett, K. M., Isaac, A., Fenlon, K. S., Doerr, M., Meghini, C., Palmer, C. L., & Jett, J. (2013). *Modeling Cultural Collections for Digital Aggregation and Exchange Environments* (text). Center for Informatics Research in Science and Scholarship. Retrieved from <https://www.ideals.illinois.edu/handle/2142/45860>

- Wickett, K. M., Renear, A. H., & Furner, J. (2011). Are collections sets? *Proceedings of the American Society for Information Science and Technology*, 48(1), 1–10.
- Wickett, K. M., Renear, A. H., & Urban, R. J. (2010). Rule categories for collection/item metadata relationships. *Proceedings of the American Society for Information Science and Technology*, 47(1), 1–10. <http://doi.org/10.1002/meet.14504701218>
- Withey, L., Cohn, S., Faran, E., Jensen, M., Kiely, G., Underwood, W., & Wilcox, B. (2011). *Sustaining scholarly publishing: New business models for university presses* (AAUP Task Force on Economic Models for Scholarly Publishing). Association of American University Presses. Retrieved from <http://www.aaupnet.org/policy-areas/future-of-scholarly-communications/task-force-on-economic-models-report>
- Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.
- Women Writers Project. (n.d.). Guidelines for Exhibit Authors. Retrieved August 28, 2017, from <http://www.wwp.northeastern.edu/research/publications/exhibits/exhibitguide.html>
- Woutersen-Windhower, S., Brandsma, R., & Universiteitsbibliotheek. (2009). Report on enhanced publications state-of-the-art. DRIVER. Retrieved from [http://dare.uva.nl/personal/pure/en/publications/report-on-enhanced-publications-stateoftheart\(a3e80802-6909-40e3-96f6-cede61a693dd\).html](http://dare.uva.nl/personal/pure/en/publications/report-on-enhanced-publications-stateoftheart(a3e80802-6909-40e3-96f6-cede61a693dd).html)
- Wrisley, D. J., Weimer, K. H., & Grossner, K. (2015). Towards a peer review of GeoHumanities projects: A community consultation. Retrieved August 28, 2017, from <http://geohum.djwrisley.com/>
- Zavalina, O. (2010). *Collection-level subject access in aggregations of digital collections: Metadata application and use*. University of Illinois at Urbana-Champaign. Retrieved from <http://hdl.handle.net/2142/16620>
- Zhang, Y., & Wildemuth, B. M. (2009). Qualitative analysis of content. In *Applications of social research methods to questions in information and library science* (pp. 308–319). Retrieved from https://www.ischool.utexas.edu/~yanz/Content_analysis.pdf

Zorich, D. M. (2008). *A Survey of Digital Humanities Centers in the United States* (CLIR pub no. 143). Council on Library and Information Resources. Retrieved from <http://www.clir.org/pubs/reports/pub143>

APPENDIX A: TYPOLOGY OF THEMATIC RESEARCH COLLECTIONS

(For full version see <http://github.com/kfenlon/collections>)

<i>Provisional type</i>	<i>Source</i>	<i>Title</i>	<i>URL</i>
<i>Definitive-source</i>	MITH	Archimedes Palimpsest	http://www.archimedespalimpsest.org/
<i>Definitive-source</i>	MITH	Dickinson Electronic Archives (DEA 1)	http://archive.emilydickinson.org/
<i>Definitive-source</i>	MITH	Dickinson Electronic Archives (DEA 2)	http://www.emilydickinson.org/
<i>Definitive-source</i>	MITH	John Milton's A Maske or Comus	http://mith.umd.edu/comus/final/index.htm
<i>Definitive-source</i>	IATH	Xiakou: Moral Landscape in a Sichuan Mountain Village	http://www.sichuanvillage.org/
<i>Definitive-source</i>	IATH	St. Gall Monastery Plan	http://www.stgallplan.org/
<i>Definitive-source</i>	IATH	The Samantabhadra Project	http://www.thlib.org/encyclopedias/literary/canons/ngb/
<i>Definitive-source</i>	IATH	The Melville Electronic Library	http://mel.hofstra.edu/
<i>Definitive-source</i>	IATH	A New Interpretive Study of the Evolution of Slavery in Hellenistic and Roman Greece	http://www2.iath.virginia.edu/meyer/
<i>Definitive-source</i>	Omeka Codex	SiteWorks: San Francisco performance 1969-85	http://siteworks.exeter.ac.uk/
<i>Definitive-source</i>	NINES	The Online Froissart	http://www.hrionline.ac.uk/onlinefroissart/
<i>Definitive-source</i>	NINES	Corpus of Middle English Prose and Verse	http://quod.lib.umich.edu/c/cme/
<i>Definitive-source</i>	NINES	Wright American Fiction	http://www.letrs.indiana.edu/cgi/t/text/text-idx?c=wright2;cc=wright2;sid=2fda8248a5b314bdf2fe730e6a46ca2d;tpl=home.tpl
<i>Definitive-source</i>	NINES	Victorian Women Writers Project	http://webapp1.dlib.indiana.edu/vwwp/welcome.do
<i>Definitive-source</i>	NINES	The Poetess Archive	http://idhmcmain.tamu.edu/poetess/
<i>Definitive-source</i>	MITH	Shelley-Godwin Archive	http://shelleygodwinarchive.org/
<i>Definitive-source</i>	MITH	The Deena Larsen	http://mith.umd.edu/larsen/
<i>Definitive-source</i>	MITH	The Shakespeare Quartos Archive	http://www.quartos.org/
<i>Definitive-source</i>	MITH	The Thomas MacGreevy Archive	http://www.macgreevy.org/index.jsp
<i>Definitive-source</i>	IATH	World of Dante	http://www.worldofdante.org/
<i>Definitive-source</i>	IATH	Leonardo DaVinci and his treatise on painting	http://www.treatiseonpainting.org/
<i>Definitive-source</i>	IATH	Traditions of exemplary women	http://www2.iath.virginia.edu/xwomen/
<i>Definitive-source</i>	IATH	William Blake Archive	http://www.blakearchive.org/
<i>Definitive-source</i>	IATH	The Complete Writings and Pictures of Dante Gabriel Rossetti [The Rossetti Archive]	http://www.rossettiarchive.org/

<i>Definitive-source</i>	Omeka Codex	A Pilgrim's Progress By Mr. Bunion	http://pilgrims-progress.richmond.edu/
<i>Definitive-source</i>	Omeka Codex	Documenting Teresa Carreño	http://documentingcarreno.org/
<i>Definitive-source</i>	Omeka Codex	MB Williams, Living and Writing the Early Years of Parks Canada	http://mbwilliams.academic-news.org/
<i>Definitive-source</i>	Omeka Codex	The Rabat Genizah Project	http://library.lclark.edu/rabatgenizahproject/
<i>Definitive-source</i>	Omeka Codex	The Travel Letters of Mrs. Kindersley	http://travel-letters.org/kindersley/
<i>Definitive-source</i>	NINES	The Swinburne Project	http://swinburnearchive.indiana.edu/swinburne/www/swinburne/
<i>Definitive-source</i>	NINES	Charles Chesnutt Archive	http://www.chesnuttarchive.org/
<i>Definitive-source</i>	NINES	The Walt Whitman Archive	http://www.whitmanarchive.org/
<i>Definitive-source</i>	NINES	The Ambrose Bierce Project	http://www.ambrosebierce.org/main.html
<i>Definitive-source</i>	NINES	The Willa Cather Archive	http://cather.unl.edu/
<i>Definitive-source</i>	NINES	From Goslar to Grasmere	http://collections.wordsworth.org.uk/gtog/home.asp?
<i>Definitive-source</i>	NINES	The Journals of the Lewis and Clark Expedition Online	http://lewisandclarkjournals.unl.edu/
<i>Definitive-source</i>	NINES	The Yellow Nineties Online	http://www.1890s.ca/Default.aspx
<i>Definitive-source</i>	NINES	The Old Bailey Online	http://www.oldbaileyonline.org/
<i>Definitive-source</i>	NINES	Romantic Circles	https://www.rc.umd.edu
<i>Definitive-source</i>	Scholars' Lab	Latvian Dainas	http://latviandainas.lib.virginia.edu/
<i>Definitive-source</i>	Scholars' Lab	Jefferson's Notes on the State of Virginia	http://jefferson-notes.herokuapp.com/
<i>Definitive-source</i>	RRCHNM	Papers of the War Department	http://wardepartmentpapers.org/
<i>Definitive-source</i>	RRCHNM	A Digital Anthology of Early Modern English Drama	http://emed.folger.edu/
<i>Definitive-source</i>	Brown	Decameron Web	http://www.brown.edu/Departments/Italian_Studies/dweb/dweb.shtml
<i>Definitive-source</i>	Brown	The Garibaldi & the Risorgimento Archive	http://library.brown.edu/cds/garibaldi/
<i>Definitive-source</i>	Brown	Luise K. Gottsched: A biography	http://cds.library.brown.edu/projects/Gottsched/
<i>Definitive-source</i>	Matrix	Quilt Index	http://www.quiltindex.org/
<i>Definitive-source</i>	Matrix	Archive of Malian Photography	http://amp.matrix.msu.edu/
<i>Evidential platform</i>	MITH	O Say Can You See: the Early Washington, D.C. Law and Family Project	earlywashingtondc.org
<i>Evidential platform</i>	IATH	Digital Yoknapatawpha	http://faulkner.iath.virginia.edu/

<i>Evidential platform</i>	IATH	Collective Biographies of Women	http://womensbios.lib.virginia.edu/
<i>Evidential platform</i>	IATH	Aquae Urbis Romae	http://www3.iath.virginia.edu/waters/
<i>Evidential platform</i>	IATH	Evolutionary Infrastructure: Boston's Back Bay Fens	http://www2.iath.virginia.edu/backbay/fenssite/html/header/home.html
<i>Evidential platform</i>	IATH	Voting Viva Voce: Unlocking the Social Logic of Past Politics	http://sociallogic.iath.virginia.edu/
<i>Evidential platform</i>	NINES	Database of the Letters of Pope Gregory VII	http://www.g7ldb.history.uni-tuebingen.de/
<i>Evidential platform</i>	IATH	Chaco Research Archive	http://www.chacoarchive.org/
<i>Evidential platform</i>	IATH	Valley of the Shadow	http://valley.lib.virginia.edu/
<i>Evidential platform</i>	Scholars' Lab	The Mind is a Metaphor	http://metaphors.iath.virginia.edu/
<i>Evidential platform</i>	Scholars' Lab	THL Places Portal	http://www.thlib.org/places/
<i>Evidential platform</i>	Brown	Florentine Renaissance Resources: Online Tratte of Office Holders 1282-1532	http://www.stg.brown.edu/projects/tratte/
<i>Evidential platform</i>	Brown	Inscriptions of Israel/Palestine	http://library.brown.edu/cds/projects/iip/info/welcome/
<i>Evidential platform</i>	Brown	Saint-Jean-des-Vignes: Archaeology, Architecture, and History of an Augustinian Monastery	http://monarch.brown.edu/
<i>Evidential platform</i>	Matrix	Slave Biographies	http://www2.matrix.msu.edu/portfolio-item/slave-biographies/
<i>Exemplar/context</i>	MITH	Black Gotham Archive	http://archive.blackgothamarchive.org/
<i>Exemplar/context</i>	MITH	Visual Accent and Dialect Archive	http://visualaccentdialectarchive.com/
<i>Exemplar/context</i>	IATH	Jefferson's University Early Life Project	http://juel.iath.virginia.edu/
<i>Exemplar/context</i>	IATH	Digital Montpellier	http://www.digitalmontpellier.org/
<i>Exemplar/context</i>	IATH	Folklore Ukraine [originally The Ukrainian Village Project]	http://www.artsrn.ualberta.ca/folkloreukraine/
<i>Exemplar/context</i>	Scalar Showcase	Performing Archive: Edward S. Curtis + 'the vanishing race'	http://scalar.usc.edu/showcase/performing-archive-edward-s-curtis-the-vanishing-race/
<i>Exemplar/context</i>	Omeka Codex	A Parcel of Ribbons	http://aparcelforribbons.co.uk/
<i>Exemplar/context</i>	Omeka Codex	A Shoebox of Norwegian Letters	http://huginn.net/shoebox/letters/
<i>Exemplar/context</i>	Omeka Codex	A Thin Ghost	http://www.thin-ghost.org/
<i>Exemplar/context</i>	Omeka Codex	American Merchant Marine Veteran's Oral History Project	http://seamenschurch-archives.org/sci-ammv
<i>Exemplar/context</i>	Omeka Codex	Bracero History Archive	http://braceroarchive.org/
<i>Exemplar/context</i>	Omeka Codex	CGP Community Stories	http://www.cgpcommunitystories.org/

<i>Exemplar/context</i>	Omeka Codex	Dante on Stamps	http://www.danteonstamps.com/
<i>Exemplar/context</i>	Omeka Codex	eBlack Champaign-Urbana: A Collaborative Portal on African-American History and Culture	http://www.eblackcu.net/
<i>Exemplar/context</i>	Omeka Codex	Environmental Design Archives Exhibitions	http://www.ced.berkeley.edu/cedarchives/exhibitions/
<i>Exemplar/context</i>	Omeka Codex	Fifteenth-Century Italian Art	http://www.quattrocentoitalia.artinterp.org/omeka/
<i>Exemplar/context</i>	Omeka Codex	Folk Horror	http://www.folkhorror.com/
<i>Exemplar/context</i>	Omeka Codex	From farms to freeways: Women's memories of Western Sydney	http://omeka.uws.edu.au/farmstofreeways/
<i>Exemplar/context</i>	Omeka Codex	Goin' North: Stories from the First Great Migration to Philadelphia	http://goinnorth.org/
<i>Exemplar/context</i>	Omeka Codex	Hurricane Digital Memory Bank	http://hurricanearchive.org/
<i>Exemplar/context</i>	Omeka Codex	I Am A Man: The Memphis Sanitation Workers Strike	http://dlxs.lib.wayne.edu/iamaman/
<i>Exemplar/context</i>	Omeka Codex	Identities: Understanding Islam in a Cross-Cultural Context	http://marb.kennesaw.edu/identities/
<i>Exemplar/context</i>	Omeka Codex	Making the History of 1989: The Fall of Communism in Eastern Europe	http://chnm.gmu.edu/1989
<i>Exemplar/context</i>	Omeka Codex	Martha Washington, a Life	http://marthawashington.us/
<i>Exemplar/context</i>	Omeka Codex	Square Dance History.org	http://squaredancehistory.org/
<i>Exemplar/context</i>	Omeka Codex	Stanislaus River Digital Archive	http://stanislausriver.org/
<i>Exemplar/context</i>	Omeka Codex	The Great Awakening: Spiritual Revival in Colonial America	http://greatawakeningdocumentary.com/
<i>Exemplar/context</i>	Omeka Codex	Voices of the Jazz Era Ballroom	http://www.jazzeravoices.org/
<i>Exemplar/context</i>	Omeka Codex	Wearing Gay History	http://wearinggayhistory.com/
<i>Exemplar/context</i>	NINES	Digital Image Archive of Medieval Music	http://www.diamm.ac.uk/
<i>Exemplar/context</i>	MITH	Early Americas Digital Archive	http://mith.umd.edu/eada/
<i>Exemplar/context</i>	IATH	The Arapesh Grammar and Digital Language Archive	http://www.arapesh.org/
<i>Exemplar/context</i>	IATH	The Countryside Transformed	http://eshore.vcdh.virginia.edu/
<i>Exemplar/context</i>	IATH	Salem Witch Trials	http://salem.lib.virginia.edu/home.html
<i>Exemplar/context</i>	IATH	Uncle Tom's Cabin and American culture	http://utc.iath.virginia.edu/

<i>Exemplar/context</i>	IATH	The Salisbury Project	http://salisbury.art.virginia.edu/
<i>Exemplar/context</i>	Omeka Codex	A Sailor's Life in the New Steel Navy	http://www.steelnavy.org/
<i>Exemplar/context</i>	Omeka Codex	Classicizing Philadelphia	https://classicizingphiladelphia.omeka.net/
<i>Exemplar/context</i>	Omeka Codex	Company G, 182nd Infantry Regiment: World War II in the Pacific	http://www.182ndinfantry.org/
<i>Exemplar/context</i>	Omeka Codex	From Kinema to Caligari: Sources	http://beforecaligari.org/sources/
<i>Exemplar/context</i>	Omeka Codex	Gothic Past	http://gothicpast.com/
<i>Exemplar/context</i>	Omeka Codex	Gulag: Many Days, Many Lives	http://gulaghhistory.org/
<i>Exemplar/context</i>	Omeka Codex	Histories of the National Mall	http://mallhistory.org/
<i>Exemplar/context</i>	Omeka Codex	Hong Kong's War Crimes Trials	http://hkwctc.lib.hku.hk/
<i>Exemplar/context</i>	Omeka Codex	Science Meets Art	http://gamma.library.temple.edu/sciencemeetsart/
<i>Exemplar/context</i>	Omeka Codex	Transatlantic Encounters: Latin American Artists in Interwar Paris	http://chnm.gmu.edu/transatlanticencounters/
<i>Exemplar/context</i>	NINES	The Vault at Pfaff's	http://digital.lib.lehigh.edu/pfaffs/about/welcome/
<i>Exemplar/context</i>	NINES	Nineteenth-Century Disability: Cultures and Contexts	http://www.nineteenthcenturydisability.org/
<i>Exemplar/context</i>	NINES	Gothic Ivories Project at the Courtauld Institute of Art	http://www.gothicivories.courtauld.ac.uk/
<i>Exemplar/context</i>	Scholars' Lab	The Fralin UVa Art Museum Numismatic Collection	http://coins.lib.virginia.edu/
<i>Exemplar/context</i>	Scholars' Lab	The Falmouth Project	http://falmouth.lib.virginia.edu/
<i>Exemplar/context</i>	Scholars' Lab	Faulkner at Virginia: An Audio Archive	http://faulkner.lib.virginia.edu/
<i>Exemplar/context</i>	Scholars' Lab	For Better for Verse	http://prosody.lib.virginia.edu/
<i>Exemplar/context</i>	Scholars' Lab	Mapping the Catalogue of Ships	http://ships.lib.virginia.edu/home
<i>Exemplar/context</i>	RRCHNM	Amboyna Conspiracy Trial	http://amboyna.org/
<i>Exemplar/context</i>	RRCHNM	A Liberian Journal	http://liberianhistory.org/
<i>Exemplar/context</i>	RRCHNM	The September 11 Digital Archive	http://911digitalarchive.org/
<i>Exemplar/context</i>	RRCHNM	Hurricane Digital Memory Bank	http://hurricanearchive.org/
<i>Exemplar/context</i>	RRCHNM	Children and Youth in History	http://chnm.gmu.edu/cyh/

<i>Exemplar/context</i>	RRCHNM	Women and World History	http://chnm.gmu.edu/wwh/
<i>Exemplar/context</i>	RRCHNM	Probing the Past: Virginia and Maryland Probate Inventories	http://chnm.gmu.edu/probateinventory/
<i>Exemplar/context</i>	RRCHNM	Imaging the French Revolution	http://chnm.gmu.edu/revolution/imaging/
<i>Exemplar/context</i>	RRCHNM	Critical Infrastructure Protection Oral History Project	http://chnm.gmu.edu/cipdigitalarchive/
<i>Exemplar/context</i>	Brown	The Whole World Was Watching: an oral history of 1968	http://www.stg.brown.edu/projects/1968/
<i>Exemplar/context</i>	Brown	A & L Tirocchi Dressmakers Project	http://tirocchi.stg.brown.edu/
<i>Exemplar/context</i>	Brown	Anne S. K. Brown Military Collection (Prints, Drawings and Watercolors)	http://dl.lib.brown.edu/askb/
<i>Exemplar/context</i>	Brown	Shadows at Dawn: A Borderlands Massacre and the Violence of History	http://www.brown.edu/Research/Aravaipa/
<i>Exemplar/context</i>	Brown	Catskills Institute	http://catskills.brown.edu/
<i>Exemplar/context</i>	Brown	The Great Kanto Earthquake of 1923	http://library.brown.edu/cds/kanto/
<i>Exemplar/context</i>	Brown	Latin American Travelogues	http://dl.lib.brown.edu/travelogues/
<i>Exemplar/context</i>	Brown	Modernist Journals Project	http://dl.lib.brown.edu/mjp/
<i>Exemplar/context</i>	Brown	Online Gazetteer of Sixteenth Century Florence	http://cds.library.brown.edu/projects/florentine_gazetteer/
<i>Exemplar/context</i>	Brown	Perry in Japan: A Visual History	http://library.brown.edu/cds/perry/
<i>Exemplar/context</i>	Brown	Romanian Love Charms	http://cds.library.brown.edu/projects/romanianCharms/
<i>Exemplar/context</i>	Brown	Underground Rhode Island	http://cds.library.brown.edu/projects/undergroundri/
<i>Exemplar/context</i>	Brown	Voyage of the Slave Ship Sally	http://cds.library.brown.edu/projects/sally/
<i>Exemplar/context</i>	Matrix	Islam and Modernity	http://aodl.org/islamicmodernity/
<i>Exemplar/context</i>	Matrix	What America Ate	http://whatamericaate.org/
<i>Exemplar/context</i>	Matrix	Pluralism and Adaptation in the Islamic Practice of Senegal and Ghana	http://aodl.org/islamicpluralism/
<i>Exemplar/context</i>	Matrix	South Africa: Overcoming Apartheid, Building Democracy	http://overcomingapartheid.msu.edu/
<i>Exemplar/context</i>	Matrix	Diversity and Tolerance in the Islam of West Africa	http://aodl.org/islamictolerance/
<i>Exemplar/context</i>	Matrix	African Activist Archive	http://africanactivist.msu.edu/

APPENDIX B: CONTENT ANALYSIS PROTOCOL

Cluster	Categories of analysis
Context	Theme; Purposes; Impact; Creators; Audience; Documentation; Provenance; Related collections; Related projects and publications; Review; Funding; Developmental stage; Host; Rights; Sustainability and preservation plans; Method
Content	Items; Diversity; Size; Narrativity; Quality; Language; Completeness; Density; Spatial coverage; Temporal coverage; Interrelatedness
Design	Data models; Navigation; Infrastructural components; Interface design; Interactivity; Interoperability; Openness of components; Identification and citation; Modes of access and acquisition; Accessibility; Flexibility

This analysis protocol is structured like a qualitative content analysis codebook: it gives a set of attributes or properties or characteristics of collections. For each attribute, it gives a brief definition (often structured as a series of questions) that helps to explain how these attributes or properties are identified and characterized for each collection. For each it also lists the sources – from the sources listed above –that make reference to this attribute or property, with selected exemplary quotations.

Note that these categories or attributes are not always named as such in the sources; these are like headings for clusters of properties identified and refined in the course of analysis / protocol-development

The sources are not all about collections specifically, but about digital humanities scholarship generally. This protocol takes an inclusive stance.

Note that while sources are often normative in their descriptions of attributes and properties, the definitions used for content analysis aimed to discard what is prescriptive or normative in the sources, and drill down to characteristics at their cores. The aim of this analysis, after all, is not to evaluate the degree to which collections adhere to evaluation standards or best practices, but to make headway on the more fundamental questions of what these things are and how they work

Context

Theme

Definition

- Subject, topic, controlling idea, conceptual core
- May be narrowly or tightly defined, even to the point of being a specific research question or objective
- Central to unity criteria and related to collection intension: items' relevance to a theme may be described in terms of Aboutness, Exemplarity, Nature, Witness (types of relevance described in Bekiari, Doerr and LeBoeuf, 2008, from Wickett, et al., 2013)

Sources

- Definition of “thematic research collection” (Unsworth, 2000; Palmer, 2004)
- “project's controlling idea” (Kuhn, Johnson and Lopez, 2010);
- “Is there a strong thesis or argument at the core of this project? Does the project clearly articulate, or some way make ‘experiential,’ this conceptual ‘core’?” (Mattern, 2012)
- “The contents are *thematic* or focused on a research theme...author-based themes, ... a literary or artistic work,...a narrowly defined literary theme... A collection theme can be an event, place, phenomenon, or any other object of study.” (Palmer, 2004)
- DC-CAP subject; AHR Prize Criteria; IDE 2017

Purposes

Definition

- Intended contributions, both explicit and (where discernable) implicit
- May be numerous, and at different levels of generality and priority
- Some purposes are entailed by the definition of “thematic research collection”: to collocate items, to support research on a theme
- Balance of primary and secondary sources
- There are a number of dimensions to this

Sources

- “Has the project fulfilled its intended purpose?” (IRH);
- “Care should be taken by the scholar to explain the unique contributions of a work, its relation to existing fields, the labor involved in its creation, the most useful ways of assessing influence and quality” (Anderson and McPherson)
- Presner, 2012; IDE, 2017; NINES/NEH Whitepaper; Rockwell, 2012; etc.

Impact

Definition

- Sources were concerned with how collections can explain and demonstrate their impacts
- This dimension falls outside of the scope of this study, but it stays in this protocol because it was common in the sources and represents a clue toward future work on the use and reuse of thematic research collections

Sources

- “Impact can be measured in many ways, including the following: support by granting agencies or foundations, number of viewers or contributors to a site and what they contribute, citations in both traditional literature and online (blogs, social media, links, and trackbacks), use or adoption of the project by other scholars and institutions, conferences and symposia featuring the project, and resonance in public and community outreach (such as museum exhibitions, impact on public policy, adoption in curricula, and so forth).” (Rockwell, 2012)

Creators

Definition

- Who created and contributed to this collection?
- What kinds of collaboration among creators was entailed?
- What disciplines, expertise, and roles were involved?

Sources

- “An entity who gathers (or gathered) the items in a collection together.” (DC-CAP)
- "What is the nature of the community that conceptualizes, organizes, and produces this scholarship? ... (Who?) Software engineers? Informatics experts? Interface designers? Archivists and librarians? Humanists and social scientists? ... What kind of conceptualization of knowledge does this collaborative set up create?" (NINES/NEH whitepaper)
- Multimedia scholarship is often produced through intense collaborations that extend across very different disciplinary traditions" (Anderson and McPherson);
- IRH; etc.

Audience

Definition

- What are the target audiences for the collection?
- How does the collection aim to serve the needs of that audience?

Sources

- “Is it directed at a clear audience? Will it serve the needs of that audience?”; “Is the project readily available for its target audience(s)?... Is it equally effective to reach all targeted audiences (for example, in multilingual/multicultural projects?) ... Are there potentially valuable unintended audiences?” (NINES/NEH Whitepaper)
- DC-CAP, etc.

Documentation

Definition

- How well documented is the collection? What aspects of the collection and its creation are documented, and for what purposes (e.g., collaborative development, sustainability and preservation, or editorial documentation)?

Sources

- IDE 2017; etc.

Provenance

Definition

- Data provenance and custodial history of items in the collection
- What changes have been made to the items in the collection, in the course of their representation?

Sources

- “Is the data provenance described and discussed? Is the scope of the data articulated? Are the authors’ manipulations of the data documented in prose or code (and thus reproducible)? Are reused data appropriately cited” (Wrisley, 2015)
- “Humanities research objects are often artifacts, created with purpose and audience, and their history and provenance is part of their identity” (Flanders and Jannidis, 2015)
- “A statement of any changes in ownership and custody of the collection that are significant for its authenticity, integrity, and interpretation” (DC-CAP)

Related collections

Definition

- Are there related collections of primary or secondary sources?
- If so, what is their relationship?
- Are there sub- or super-collections?

Sources

- “A second collection that is associated with the current collection.” (DC-CAP)

Related projects and publications

Definitions

- What projects are related to this collection?
- What publications are associated with the collection?
- Both aspects are concerned with the collection’s relationship to other scholarship.

Sources

- “How does it relate to other printed or digital resources, to its predecessors or to similar projects?” (IDE 2017)
- “Does the project fully engage with current scholarship in the field?” (DHCOMMONS)
- “Is the project linked to or affiliated with other projects? ... Do other projects acknowledge this project?” (NINES/NEH Whitepaper)

Review

Definitions

- Has the collection been reviewed or evaluated in any formal way?

Sources

- “Have there been any expert consultations? Has this been shown to others for expert opinion?” (Rockwell)

Funding

Definition

- How has the collection been supported?

- How has funding informed the growth or development of the collection?

Sources

- “Resources should be ‘badged’ according to the funding bodies’ requirements. Approved logos of funding bodies and lead institutions where applicable should be displayed at the start (access) page of the resource” (IRH);
- “competitive funding decisions like the allocation of a grant should be considered as an alternative form of review” (Rockwell, 2012)

Developmental stage

Definitions

- At what stage of development is the collection?
- Is the collection in active development, subject to future change, or in stasis? Is the collection considered “complete” or “finished”?

Sources

- “It should be made clear when a resource was first made available and when it was last updated. // If a resource is ‘in progress’ it should have record-level information about the currency of particular records” (IRH);
- “At what ‘stage’ is the project in its current form? Is it considered ‘complete’ by the creators, or will it continue in new iterations, perhaps through spin-off projects and further development?” (Presner, 2012)

Host

Definition

- How is the collection hosted? Who is responsible for “publishing” the collection?
- What are the implications of the host or imprint for the collection?

Sources

- “How is the project hosted? Through a university server? A commercial host? A non-profit organization? Is there evidence of ongoing commitment to support of the project at the level of hosting? Is there similar evidence of ongoing support from project personnel?” (DHCommons)
- “The nature of the organization mounting a web resource is one sign of the background of a digital project. ... Evaluators can ask about the nature of the organization that hosts a project as the act of hosting or mirroring (providing a second ‘mirror’ site on another server) is often a recognition of the worth of the project.” (Rockwell, 2012)

Rights

Definitions

- What are the rights held over items in the collection, the software underlying the collection, and other components?

Sources

- “Information about rights held in and over the resource.” (DC-CAP)

Sustainability or preservation plan

Definition

- Is there a plan for the project’s ongoing maintenance or ultimate preservation?

Sources

- “Is there a preservation and maintenance plan for the interface, software, and associated databases (multiple copies, mirror sites, collaboration with data archives, etc.)? Is the project fully exportable/transferrable?” (DHCommons)
- “What are the [scholarly digital edition]’s prospects for long term use? Is the edition complete or does it promise further modifications and additions? Is there institutional support for the curation and sustainment of the [scholarly digital edition]? Is the basic data archived? Is there a plan to provide continuous access to the presentation?” (IDE);
- “How will the project ‘live’ and be accessible in the future, and what sort of infrastructure will be necessary to support it? Here, project specific needs and institutional obligations come together at the highest levels” (Presner, 2012)
- “How would you assess its reliability and long-term value?” (IRH)

Method

Definition

- What methods have been employed in the creation of the collection? And what implications do they have for the use and interpretation of the collection?

Sources

- “Which editorial school does the [scholarly digital edition] follow? Which methodological approach does it take? Does it apply e.g. a materialistic or an idealistic / platonic understanding of text? Is it focussing on ‘works’ or on ‘documents’? How does it assess the textual tradition: Are there preferred manuscripts or are all documents considered to be of equal value?” (IDE 2014)
- “How does the project build up the dataset?” (IDE 2017)
- “Do the digital methods employed offer unique insights into the project’s key questions?” (DHCOMMONS)
- “A method by which items are added to a collection.” (DC-CAP Accrual Method)

Content

Items

Definition

- What constitutes an item in this collection? Often the conceptual unit that a user might identify as an item of interest may not be located in any particular digital object, but in the interrelation of many different digital objects and processes.

- How we can characterize discrete items in the collection, holding that term loosely, and focusing on the conceptual?
- What are the main units of analysis and description?
- What units are returned by searching and browsing mechanisms?
- Primary sources are usually the main “items” within these collections; how are these represented and mediated by affordances or other components of collections?

Sources

- Definition of “collection” (Unsworth, 2000; Palmer, 2004)
- DC-CAP
- AHR Prize Criteria

Diversity

Definition

- Heterogeneity of items, in terms of formats, media, types, etc.
- Balance of primary and secondary sources
- There are a number of dimensions to this

Sources

- “the content is *heterogeneous* in the mix of primary, secondary, and tertiary materials provided” (Palmer, 2004)
- “the use of multiple modes of communication, ... Still and moving images, interactive maps, audio, and other media must be integrated with one another, and especially with the core of textual argumentation” (AHR Prize Criteria)
- “Does it effectively “triangulate” a variety of sources and make use of a variety of media formats?” (Mattern)
- JAH Digital History Reviews; IML; Kuhn, Johnson & Lopez; DHCommons; NINES/NEH Whitepaper
- “Think about opportunities to incorporate multimedia in ways that take advantage of your digital work. Database searching, dynamic user-directed displays, audio and video represent scholarly opportunities that go beyond traditional print scholarship.”; “Does the project’s form suit its concept and content? “Do structural and formal elements of the project reinforce the conceptual core in a productive way?” (University of Nebraska)

Size

Definition

- The size or extent of a collection
- Usually measured by item count, though this is not a simple measure: e.g., do you count the encoded manuscript pages individually, or do you only count whole works?
- Extent is a common theme in descriptions of collections, but there is some disagreement over whether collections should or must be of a certain size; e.g., Thomas characterizes them as “sprawling”, but collections surveyed for this study ranged from just a few items to many thousands of items. The scope of a collection seems more related to the capaciousness of a theme, or the extents of available collections of sources.

- Some characterize extent in terms of intellectual contribution or labor entailed (e.g., WWP Guidelines, below)

Sources

- “The size of the collection.” (DC-CAP);
- “Exhibits of almost any length may be submitted, but a practical range might be from at least one substantial paragraph to a book chapter. On the shorter end, the paragraph would need to be very substantive and suggestive; on the longer end, the piece would truly have to earn the space by holding the reader's attention.” (WWP Guidelines for exhibit authors)

Narrativity

Definition

- To what extent does the collection employ narrative description or discussion, to express interpretation, analysis, or contextual information?
- To what extent or in what way are items framed with, embedded in, or complemented by linear, narrative scholarly contributions?
- Does the collection include or link to original or unoriginal, narrative secondary sources?
- What is the relationship between the discursive elements and the intended contribution of the collection?
- Tone or narrative style

Sources

- AHR Prize Criteria; IRH recommendations; NINES/NEH Whitepaper; Kairos source

Quality

Definition

- Quality is a common attribute or characteristic discussed in the sources – largely pertaining to standards-compliance and following best-practice guidelines for digitization quality.
- Where does the collection make its investments into quality? Items? Page images? Encoding? How does this affect the meaning or evidential value of the collection?

Sources

- IRH; Rockwell, 2012; Mandell, 2012; IDE, 2014; NINES/NEH Whitepaper

Language

Definition

- What languages are employed in the collection?

Sources

- “A language of the items in the collection.” (DC-CAP)

Completeness

Definition

- The ideal of completeness toward which a collection aims to grow
- Definitiveness (exhaustiveness), exemplarity (representativeness), or evidential sufficiency (defined variously in accordance with research objectives)

Sources

- “Is relevant content missing? Is any omission explained and/or justified?” (IDE 2014)
- “claims to be representative for a specific subject domain” or “functions as a reference for that domain” (IDE, 2017)

Density

Definition

- Topical homogeneity, or thematic coherence
- How closely related are items in terms of their subjects?
- Particularly relevant in collections whose selection criteria depends on the “aboutness” of items related to theme (see “Theme,” above)
- This attribute does not apply in collections built around certain kinds of themes and it should not be taken as a universal value. Some collections will derive value from being diverse in their subjects, but tightly circumscribed in some other dimension (e.g., a collection of diverse writings from a narrow time period or geographic area).
- This is closely related to “theme” and has to do with how theme is realized by a collection of items; may be determined by a collection’s theme and selection criteria

Sources

- “A large number of subject-focused collections indicates high density. Lower density is associated with more subject-inclusive collections.” (Palmer et al., 2010)
- “creators select materials in a highly focused and deliberate manner, creating dense, interrelated collections” (Palmer, 2004)

Spatial coverage

Definitions

- Spatial scope of the collection
- Geographic areas, locations of topic or origin, or spatial coverage of the items in the collection
- May be determined by a collection’s theme and selection criteria

Sources

- “An indicator of the spatial scope of the collection.” (DC-CAP)

Temporal coverage

Definitions

- Temporal scope of the collection
- Time periods or dates of the items in the collection
- May be determined by a collection's theme and selection criteria

Sources

- “An indicator of the temporal scope of the collection.” (DC-CAP)

Interrelatedness

Definitions

- Beyond “isGatheredInto”, what other relationships obtain between collections and items?
- Between items and other objects, entities, resources, etc.?
- How are relationships expressed or implemented?

Sources

- “how do the boundedness and internal cohesion of a collection help to define its intellectual purpose?” (Flanders, 2014)
- “creators select materials in a highly focused and deliberate manner, creating dense, interrelated collections” (Palmer, 2004)

Design

Data models

Definition

- How are items represented and how are those representations determined by underlying data models?
- What are the implications of the data models chosen to represent items? Especially, how are data models purposively employed to create coherence between how items are represented and the overarching purposes of the collection?
- For example, metadata schemas, encoding or markup structures, and enrichments to sources in the form of annotations or other structures
- While it is often difficult to unravel data models from other technical aspects of collections, such as navigational features, other infrastructural components, and interface design, the distinctions may sometimes be revelatory

Sources

- “What decisions and choices have been made regarding the representation of the materials?” (NINES/NEH whitepaper)

- “Is there a clear statement of the standards that have been used, and an explanation of their benefits and/or limitations? Have the data been well constructed?” (IRH);
- “software and hardware only put into effect the models structured into their design” (Drucker, 2009)
- “How is the editorial method technically implemented? What data model is applied?” (IDE 2014)

Navigation

Definition

- How may a collection be navigated?
- How are search and browse supported and how may users navigate among items and other resources?
- How are classification and faceting schemes employed to organize search and browse across the collection?
- What do navigational features render more visible and most prominent?

Sources

- Rockwell (2012); LAIRAH checklist (Warwick et al., 2007); ARC peer review guidelines; IRH recommendations (Bates et al., 2006); IDE 2014 (search); etc.

Infrastructural components

Definition

- What other infrastructural components of collections, beyond data models, function to enable and constrain collections’ representations, functions, and uses?
- Discrete technological components, ranging from database systems and search algorithms to tools used for data manipulation to support advanced functionalities

Sources

- “Does the project ‘exhibit an understanding of the affordances of the tools used,’ and does it exploit those affordances as best possible — and perhaps acknowledge and creatively ‘work around’ known limitations?”; “tools...instantiate hermeneutical positions about what questions are important” (MLA wiki?), cited in Anderson and McPherson (2011)
- “The ‘digital’ tools and modalities of the work must contribute substantially to the argument presented by the author.” (AHR)
- “How does the work advance an argument through both the content and the way the content is presented? How is the design of the platform an argument?” (IRH)

Interface design

Definition

- This study specifically considers two aspects of the interface design of collections:
- how they organize elements on the screen, and
- what rhetorical functions the appearance and aesthetic of the collection can be discerned to serve.

- Because interface design is often the least-documented aspect of the technical design of a digital resource, and to avoid reading too much into design choices, this study focuses on what can be empirically determined about a collection's interface, such as what imagery a collection foregrounds, what elements it makes most and least visible, and how it interrelates elements on the screen.

Sources

- “How do the design and content elements of the project interact and integrate with one another?” (Wrisley, et al.)
- “solid design principles so that the resource promotes rather than deters thinking” (18th Connect...)
- “scholars are free to organize and design their materials as they judge best, given the purposes and goals of the project” (JAH)
- “As Kress (2010) has said, 'Design is the servant of rhetoric'—or, to put it differently: the political and social interests of the rhetor are the generative origin and shaping influence for the semiotic arrangements of the designer” (Ball, 2012)

Interactivity

Definition

- How does the collection facilitate or enable interaction between user and collection, beyond searching, browsing, and reading/viewing?
- Some collections create interactivity through built features, which allow different kinds of use of items.
- Some facilitate interactivity not only with the collection itself, but with other users of the collection, turning the collection into a potential hub for collaboration or discourse, for example by adding links to social media or forums, or adding annotation, commenting, or feedback mechanisms.
- Other collections enable interactivity not by building additional uses or functions into the site, but by removing barriers to unanticipated use of collection contents and code, e.g., by opening their underlying data to use.

Sources

- Anderson and McPherson, 2011; IDE; Palmer, 2004

Interoperability

Definition

- How does the collection interoperate or anticipate interoperating with other resources, other collections, and external tools?

Sources

- “Attempts to render the work interoperable with other digital resources.” (Mandell, 2012 citing MLA Guidelines)
- “One indication of how a digital work participates in the conversation of the humanities is how it links to other projects and how in turn, it is described and linked to by others.” (Rockwell, 2012);

- “Does the project exploit the ‘repurpose-ability’ of data? Does it pull in, and effectively re-contextualize, data from other projects?” (Mattern, 2012)

Openness of components

Definition

- How readily and freely can data and other components can be accessed, used, reused, etc.?
- This is not a flat value for digital collections: despite the commonness of “openness” as a desired property of collections, Nowviskie among others has noted that this property is not appropriate for all kinds of collections (cite)
- Despite its potential normativity, the prevalence of “openness” in the literature suggests that it deserves representation
- How do the different ways that a collection may manifest “openness” reflect its distinctive purposes, and differently affect its immediate and long-term use?

Sources

- Anderson and McPherson, 2011; IRH; 18th Connect and NINES; etc.

Identification and citation

Definition

- The ideal of completeness toward which a collection aims to grow
- Definitiveness (exhaustiveness), exemplarity (representativeness), or evidential sufficiency (defined variously in accordance with research objectives)

Sources

- “Is relevant content missing? Is any omission explained and/or justified?” (IDE 2014)
- “claims to be representative for a specific subject domain” or “functions as a reference for that domain” (IDE, 2017)

Modes of access and acquisition

Definition

- At what level are primary sources identified; what’s the unit of identification?
- Is this any reflection of intended contribution?
- How are resources identified?

Sources

- “Are there technical interfaces like OAI-PMH, REST, APIs etc., which allow the reuse of the data of the [scholarly digital edition] in other contexts? Can you harvest or download the data? Can you use the data with other tools useful for this kind of content? Can you integrate the content in other systems, e.g. aggregating content from several sources?...Is the basic or

underlying data of the edition accessible (e.g. in XML) and if so, how? Is it provided for each single object and/or for the whole [scholarly digital edition]? Is the access part of the [scholarly digital edition]'s user interface or part of an external repository? If you cannot access the basic data, is a justification provided?" (IDE 2014)

Accessibility

Definitions

- Is the collection designed to be accessible to users with disabilities?

Sources

- "Resources should be compliant with general accessibility standards and CENDAR requirements. // Resources' claims to technical and accessibility standards adherence should match actuality. // In case of non-adherence to particular standards, resources should outline in their documentation the reasons why those standards are not being followed. // Resources should have flexible screen widths. // Resources should not use pop-up windows unless necessary" (IHR)

Flexibility

Definitions

- How is the collection designed to be flexible, extensible, mobile, adaptable, repurposeable, remixable, dynamic, adaptive?

Sources

- "How flexible is the resource? Has it been designed in such a way that future developments can be easily incorporated" (IRH)
- "Can new materials be added effectively? ... Can problems be identified and fixed easily? ...Can it be easily migrated to new platforms?" (NINES/NEH Whitepaper)

APPENDIX C: INTERVIEW PROTOCOL

Interview Part 1. For participants engaged in development of DH projects and digital collections:

1. Tell me about your role at your institution.
Prompts:
 - a. What is your title?
 - b. How long have you been in this role?
 - c. How do you and your unit relate to other units on campus, e.g., the University Library, scholarly communications office, or digital humanities center?
2. Tell me about your experience with digital research collections, especially those created by scholars.
3. How do you or how does your unit find, recruit, or select digital projects to develop?
Prompts:
 - a. What criteria do you use to select among digital projects?
 - b. How proactive are you, in the recruitment of projects? Do authors come to you, or do you produce calls for proposals, etc.?
4. In a recent project that resulted in a digital collection, how did development proceed?
Prompts:
 - a. Describe the digital collection/project.
 - b. Walk me through the process of creating a digital collection, or helping a scholar to develop a digital collection.
 - c. What role did you play, and what roles did your center/library play?
 - d. What factored into the selection of content for the collection, and its representation?
 - e. What factored into decisions about (or how are choices made among) underlying architectures or data models?
 - i. What were the priorities, in deciding how to undergird the collection?
 - ii. Was longevity or future maintenance/preservation a consideration?
 - f. What factors into decisions about (or how are choices made among) forms and functionalities the digital collection will have?
 - g. How long was this project?
 - h. Is this process typical of digital humanities projects at your institution?
 - i. Can you provide any contrastive examples?
5. What kinds of services or support do you provide to scholars/authors in the process of development of digital collections?
Prompts:
 - a. In your experience, what are common challenges that confront authors or creators of digital projects and publications, and for which you provide support?
 - b. Examples of common publishing services:
 - i. Organizational and conceptual input on content
 - ii. Representation of content, both structural and design-related representation
 - iii. Navigating third-party permissions for content
 - iv. Navigating other copyright concerns
 - v. Technological instruction
 - vi. Website, database, tool, or other resource development
 - vii. Web hosting
 - viii. Funding

- ix. Project planning
 - c. What kinds of services or support do scholars seek that you are *unable* to provide?
 - d. Do these collections undergo editorial review, peer review or any other sort of assessment or evaluation either in the course of or after development?
 - i. If so, what sorts of review, and how are they conducted?
 - ii. What are benchmarks for success, or how do you or the scholars you work with evaluate the outcomes?
 - iii. Do you make any assessment or track usage of digital collections or DH projects?
 - e. Are these digital collections marketed or promoted in any way, and if so, how?
 - f. In the course of development, do you make provisions for long-term maintenance, preservation, or archiving of these projects or collections, after project end?
 - i. If so, of what sort?
 - 6. How do these projects relate to your institution's overarching programs related to scholarly publishing or scholarly communication?
- Prompts:
- a. If your university has a press, are they involved in these kinds of digital projects?
 - b. To your knowledge, does your university library support scholarly publishing?
 - i. In what ways?
 - ii. Does your library's scholarly publishing program relate to these digital humanities efforts, and if so, how?
 - c. How do you think digital humanities projects at your institution fit into or participate in scholarly communication or publishing generally?
7. Are you or is your unit actively involved in ongoing or long-term management, maintenance, preservation, or archiving of scholarly digital collections?
- Prompts:
- a. [If yes, proceed to interview part 2]
 - b. If not, is anyone or any unit responsible for ongoing maintenance of these resources? [If yes, request recommendation for relevant contact]
 - c. If no one is overseeing the ongoing maintenance of these resources, can you offer any thoughts on why, or what challenges prevent their ongoing management or preservation?
 - d. Can you think of anyone else at your institution who might speak with authority on alternative publications, who I should talk to?

Interview Part 2. For participants engaged in ongoing or long-term management, maintenance, preservation, or archiving of scholarly digital collections:

- 1. [If interviewing a different participant at the institution than in interview part 1, repeat preliminary questions 1 and 2 from above]
- 2. How is your unit involved with the ongoing management, maintenance, or preservation of scholarly digital collections?

Prompts:

- a. How would you describe the nature of your unit's involvement?
 - i. Hosting vs. active maintenance vs. preservation or archiving?
- b. What is the extent of work that your unit does in this area?
- c. What sorts of provisions are made, by your unit or any other, for the active and ongoing management of scholarly collections throughout their lifecycle of interest and usefulness to scholarship?

- d. Do you make any ongoing assessment of the use of these collections?
 - e. Are collections “selected” for this treatment, and if so, how?
- 3. Are there collections that are planned to be in active development for the foreseeable future, or indefinitely incomplete?
 - a. How are these handled?
- 4. Are these collections integrated into any systems or tools for discovery and access?

Prompts:

- a. Within or outside of the library?
 - b. Are they given any kind of structured description?
 - c. Do you see a need for integration of these collections into systems of discovery?
- 5. [If not at/affiliated with the library:] Is the library involved with the ongoing maintenance of scholarly digital collections? If so, how?
- 6. To your knowledge, are any open-access resources “collected” (in any sense), made discoverable and accessible, or managed in an ongoing way by the library? If so, how?
 - a. What do you see as obstacles to library “collection” of open-access materials?
 - b. Do you see potential benefits?
- 7. What are the greatest challenges you see to keeping these kinds of collections discoverable and accessible over the long term?
- 8. We have discussed how things are. In your view, what would be the optimal strategies for ongoing management of scholarly collections?

Prompts:

- a. Who would bear responsibility for ongoing management?
- b. What advantages and disadvantages do you see to library involvement?
- c. What level of management is needed or optimal?
- d. [If no one is responsible for ongoing maintenance:] Why do you think this is the case?